

2016

Network-coded cooperative communications with multiple relay nodes: Achievable rate and network optimization

Sushant Sharma

Virginia Polytechnic Institute and State University

Yi Shi

Virginia Polytechnic Institute and State University

Y. Thomas Hou

Virginia Polytechnic Institute and State University, thou@vt.edu

Sastry Kompella

U.S. Naval Research Laboratory

Scott F. Midkiff

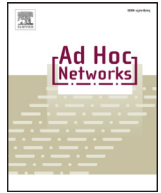
Virginia Polytechnic Institute and State University

Follow this and additional works at: <http://digitalcommons.unl.edu/usnavyresearch>

Sharma, Sushant; Shi, Yi; Hou, Y. Thomas; Kompella, Sastry; and Midkiff, Scott F., "Network-coded cooperative communications with multiple relay nodes: Achievable rate and network optimization" (2016). *U.S. Navy Research*. 119.

<http://digitalcommons.unl.edu/usnavyresearch/119>

This Article is brought to you for free and open access by the U.S. Department of Defense at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in U.S. Navy Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Network-coded cooperative communications with multiple relay nodes: Achievable rate and network optimization

Sushant Sharma^a, Yi Shi^a, Y. Thomas Hou^{a,*}, Sastry Kompella^b, Scott F. Midkiff^a

^a Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

^b U.S. Naval Research Laboratory, Washington, D.C., USA

ARTICLE INFO

Article history:

Received 28 January 2016

Revised 8 September 2016

Accepted 13 September 2016

Available online 14 September 2016

Keywords:

Cooperative communications

Network coding

Relay nodes

ABSTRACT

Network-coded cooperative communications (NC-CC) refers to the use of network coding (NC) in cooperative communications (CC). Prior studies have shown that NC has the potential to improve the performance of CC when there are multiple sessions in the wireless network. These studies were done for the case when multiple sessions are sharing a single relay node. However, how NC-CC behaves when multiple relay nodes are employed remains an open problem. In this paper, we explore this problem by analyzing the achievable rate of each session in this setting. We develop closed form formulas for the mutual information and the achievable data rate for each session. We show that prior results for a single relay is a special case of our result. Based on these findings, we then study a network optimization problem that requires joint optimization of session grouping, relay node grouping, and matching of session/relay groups. We show that this problem is NP-hard, and present a polynomial time heuristic algorithm to solve this problem. Using simulation results, we show this algorithm is highly competitive and can produce results that are near to optimality.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Cooperative communications (CC) is an important technique to improve the performance of a wireless network [11]. Unlike MIMO, which requires the use of multiple antennas at each node, CC only employs a single antenna at each node and exploits diversity by cooperating with antennas on other nodes in the network. CC schemes can be amplify-and-forward (AF) or decode-and-forward (DF) [15]. Under AF, the relay simply amplifies its received signal while under DF, the relay decodes its received signal and then encodes data again before forwarding to the destination. There has been extensive research at the physical layer that exploits cooperations among distributed antennas [1,6,13].

Recently, it was found that network coding (NC) can further improve the performance of CC by combining data streams at a relay node [4,18–25,29,33–36]. This application of NC in CC is called network-coded CC (or NC-CC). NC has been shown to improve the outage probability [4,19], packet error rates [33], and data rates [25,36] for CC. NC schemes can be either digital or analog [10], depends on whether network coding is done on digital signals or analog signals. Most of these studies were done

for the case when multiple sessions are sharing a single relay node [4,19,25,33,36] or multiple sessions have the same destination [18,20–24,34,35]. Topakkaya and Wang [29] considered the scenario of multiple source-destination pairs using multiple relays, designed a network coding scheme, and analyzed its performance. In this paper, we will design another network coding scheme with less number of time slots in a frame and thus may achieve larger rate.

In this paper, we study NC-CC when there are multiple relay nodes. Our goal is two-fold.

First, we aim to develop closed form formulas for the mutual information and the achievable data rate for each session. We consider the case of analog network coding (ANC) [10] and AF CC [15] at each relay node. Through an in-depth analysis, we derive the mathematical equations for mutual information and achievable data rate for each session and show that prior results for a single relay is a special case of our result. This finding offers an important building block on the theory of NC-CC.

Second, we investigate the following important problems jointly in a multi-user network under NC-CC: (i) how to put sessions into different groups; (ii) how to put relay nodes into different groups; and (iii) how to match the session groups with relay groups under NC-CC. Specifically, we study a network optimization problem with the goal of maximizing the sum of weighted rates of all sessions. This optimization problem requires a joint optimization of all three

* Corresponding author. Fax: +1 540 231 8292.
E-mail address: thou@vt.edu (Y. Thomas Hou).

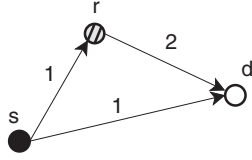


Fig. 1. A three-node relay channel for CC.

Table 1
Notation.

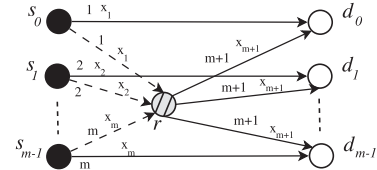
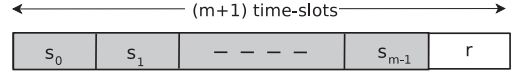
Symbol	Definition
NC-CC	Network Coded Cooperative Communications.
W	Total bandwidth available in the network
SNR_{uv}	The signal to noise ratio between nodes u and v
α_r	Amplification factor at relay node r
σ_v^2	Variance of background noise at node v
$\sigma_{z_v}^2$	Variance of ANC noise at node v
h_{uv}	Effect of path-loss, shadowing, and fading from node u to node v
z_v	White Gaussian background noise at node v
z_v^{ANC}	ANC noise at node v
w_i	Weight assigned to session (s_i, d_i)
x_i	Signal transmitted by source s_i
y_{uv}	Signal transmitted by node u and received by node v
$y_{\mathcal{R}d_i}$	Signal transmitted by relay group \mathcal{R} and received by node d_i
$y_{\mathcal{S}d_i}$	Signal received by destination d_i that was originally transmitted by session group \mathcal{S} and then retransmitted by relay group \mathcal{R}
h_{uv}	Effect of path-loss, shadowing, and fading from node u to node v
P_u	Transmission power at node u
\mathcal{R}	The set of all relay nodes in the network
\mathcal{S}	The set of all source nodes in the network
\mathcal{R}_j	A group of relay nodes
$\mathcal{S}_{s_i}^{\mathcal{R}_j}$	Group of sessions containing s_i and using relay group \mathcal{R}_j for NC-CC

components. We show that this problem is NP-hard. Subsequently, we develop a highly competitive and efficient algorithm to solve this problem.

The remainder of this paper is organized as follows. In Section 2, we review state-of-the-art results on NC-CC when only a single relay node is employed. In Section 3, we study NC-CC with multiple sessions and multiple relay nodes. We develop formulas for the mutual information and achievable data rate of each session. In Section 4, we describe the session/relay grouping and matching problem in detail. We also show that this problem is NP-hard. In Section 5, we present an algorithm to this problem. Section 6 presents numerical results to demonstrate the performance and efficiency of the proposed algorithm. In Section 7, we discuss related work, and Section 8 concludes this paper. Table 1 lists all notation used in this paper.

2. Background

As a simple example, Fig. 1 shows a three-node relay channel for CC. In this example, source node s intends to transmit to destination node d and will exploit relay node r for possible performance improvement. Assume the time frame for transmission is divided into two time slots. We show time slot for each link in Fig. 1. That is, in the first time slot, s transmits to d , and is overheard by the relay node r . Relay node r then amplifies the received signal, and then retransmits the amplified signal in the second time slot. The destination node d can now combine the two copies of the same signal coming from two different paths. This cooperative relay channel in Fig. 1 can be treated as a single-input two-output complex Gaussian channel [15] and the achievable data

(a) m source nodes sharing one relay node.

(b) Structure of a time frame.

Fig. 2. NC-CC with a single relay node.

rate between s and d in this channel is given as follows:

$$C_{\text{CC}}(s, r, d) = \frac{W}{2} \log_2 \left(1 + \text{SNR}_{sd} + \frac{\text{SNR}_{sr} \cdot \text{SNR}_{rd}}{1 + \text{SNR}_{sr} + \text{SNR}_{rd}} \right), \quad (1)$$

where SNR_{uv} is the signal-to-noise ratio at the receiver v when node u transmits, and is given by $\text{SNR}_{uv} = \frac{|h_{uv}|^2 P_u}{\sigma_v^2}$, σ_v^2 is the variance of background noise at node v , h_{uv} is the gain of the channel from node u to node v , P_u is the power at which node u transmits signals, and W is the channel bandwidth.

As for comparison, when CC is not used, i.e., s transmits to d without using r (so-called direct transmission), the achievable rate from s to d is given as:

$$C_D(s, \emptyset, d) = W \log_2 (1 + \text{SNR}_{sd}), \quad (2)$$

where \emptyset denotes that no relay node is used. It has been shown in [15] and [28] that CC has the potential to increase achievable rate over direct transmission, depending on the location of relay node and its channel statistics.

When there are m source-destination sessions sharing a single relay node (as shown in Fig. 2), one can employ NC to combine the signals from the m sources at the relay node and then forward the combined signal to all the destination nodes, i.e., NC-CC. Here, a time frame is divided into $(m+1)$ time slots (see Fig. 2b), with time slot i , $i \leq m$, being used for transmission by source node s_{i-1} . Again, each of such transmissions is received by its corresponding destination node and overheard by the relay node. The relay node will then apply NC to combine all the m received analog signals. The combined signal x_m is then amplified and forwarded by the relay node to all destination nodes in the $(m+1)$ th time slot. The achievable rate for a session under NC-CC is given by Sharma et al. [25]:

$$C_{\text{NC-CC}}(s_i, \mathcal{S}, r, d_i) = \frac{W}{m+1} \log_2 \left(1 + \text{SNR}_{s_i d_i} + \frac{\text{SNR}_{s_i r} \text{SNR}_{r d_i}}{|\mathcal{S}_r| \frac{\sigma_{z_{d_i}}^2}{\sigma_{d_i}^2} + \text{SNR}_{r d_i} + \frac{\sigma_{z_{d_i}}^2}{\sigma_{d_i}^2} \sum_{s_j \in \mathcal{S}} \text{SNR}_{s_j r}} \right), \quad (3)$$

where $\mathcal{S} = \{s_0, s_1, \dots, s_{m-1}\}$ is the set of all the source nodes, and $\sigma_{z_{d_i}}^2$ is the noise at destination d_i due to ANC, and is given by

$$\sigma_{z_{d_i}}^2 = \sigma_{d_i}^2 + (|\mathcal{S}| - 1) (\alpha_r h_{r d_i})^2 \sigma_r^2 + \sigma_{d_i}^2 \sum_{s_j \in \mathcal{S}, s_j \neq s_i} \left(\frac{\alpha_r h_{s_j r} h_{r d_i}}{h_{s_j d_i}} \right)^2, \quad (4)$$

where α_r is the amplification factor for AF CC at relay node r and is given by

$$\alpha_r^2 = \frac{P_r}{|\mathcal{S}| \sigma_r^2 + \sum_{s_i \in \mathcal{S}} P_{s_i} |h_{s_i r}|^2}. \quad (5)$$

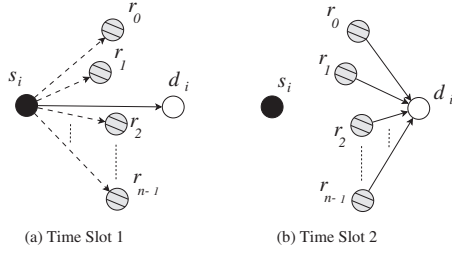


Fig. 3. CC with multiple relays.

3. NC-CC With multiple relay nodes

The NC-CC model in Fig. 2 employs only a single relay node. In this section, we study the general NC-CC model with multiple relay nodes, and derive mutual information and achievable data rate for each session. Our description of NC-CC model with multiple relay nodes is divided into two parts. First, we generalize the three-node relay channel model in Fig. 1 with multiple relay nodes. Building upon this result, we then investigate the general model where a group of sessions share multiple relay nodes in NC-CC. Note that we aim to develop closed form formulas for the achievable data rate for multiple sessions using multiple relays. The coding scheme to achieve this rate can be designed by extending the schemes in [2,3,32], which can be a separate work.

3.1. Generalizing the three-node cooperative relay channel with multiple relays

We begin by extending the three-node model in Fig. 1, and consider a model where multiple relay nodes are used by the single session for AF CC. Due to single session, NC is not needed in this model. Fig. 3 shows an example of this model. In Fig. 3, the single session is denoted by (s_i, d_i) , and the group of n relay nodes denoted by $\mathcal{R} = \{r_0, \dots, r_{n-1}\}$.

The transmission from s_i to d_i is divided into two time-slots. In the first time-slot, source node s_i transmits a signal x . The signal is received by the destination node d_i , and the relay nodes in \mathcal{R} . The signals received by the destination node d_i , and some relay node $r_j \in \mathcal{R}$ can be written as

$$y_{s_i d_i} = h_{s_i d_i} x + z_{d_i} \quad (6)$$

$$y_{s_i r_j} = h_{s_i r_j} x + z_{r_j}. \quad (7)$$

After receiving the signal, all the relay nodes in \mathcal{R} will amplify and simultaneously transmit the amplified version of the received signal to the destination node. By appropriately adjusting carrier phases, carrier frequencies and symbol timings at the transmitting nodes, signals at d_i can be combined coherently [30]. This coherent combination of signals at d_i results in a high signal-to-noise ratio (SNR) at d_i . The signal received by d_i in the second time slot can be written as:

$$\begin{aligned} y_{\mathcal{R} d_i} &= \sum_{r_j \in \mathcal{R}} (h_{r_j d_i} \alpha_{r_j}) y_{s_i r_j} + z_{d_i} \\ &= \sum_{r_j \in \mathcal{R}} [h_{r_j d_i} \alpha_{r_j} (h_{s_i r_j} x + z_{r_j})] + z_{d_i}, \end{aligned} \quad (8)$$

where z_{d_i} is the background noise at node d_i . The variable α_{r_j} is the amplification factor at r_j , and is given as:

$$\alpha_{r_j}^2 = \frac{P_{r_j}}{\sigma_{r_j}^2 + P_{s_i} |h_{s_i r_j}|^2}.$$

Eqs. (6) and (8) can be written in the following compact matrix form

$$\mathbf{Y} = \mathbf{H}\mathbf{x} + \mathbf{B}\mathbf{Z},$$

where

$$\mathbf{Y} = \begin{bmatrix} y_{s_i d_i} \\ y_{\mathcal{R} d_i} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} h_{s_i d_i} \\ \sum_{r_j \in \mathcal{R}} h_{s_i r_j} \alpha_{r_j} h_{r_j d_i} \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 0 \\ \alpha_{r_0} h_{r_0 d_i} & \alpha_{r_1} h_{r_1 d_i} & \dots & \alpha_{r_{n-1}} h_{r_{n-1} d_i} & 0 & 1 \end{bmatrix},$$

$$\text{and } \mathbf{Z} = \begin{bmatrix} z_{r_0} \\ z_{r_1} \\ \vdots \\ z_{r_{n-1}} \\ z_{d_i} \\ z_{d_i} \end{bmatrix}.$$

We can model the above channel as a one-input two-output complex Gaussian channel [15]. The mutual information between s_i and d_i for the above channel can be written as:

$$I_{CC}(s_i, \mathcal{R}, d_i) = \log \det \left\{ \mathbf{I} + (P_{s_i} \mathbf{H} \mathbf{H}^\dagger) (\mathbf{B} \mathbf{E}[\mathbf{Z} \mathbf{Z}^\dagger] \mathbf{B}^\dagger)^{-1} \right\}, \quad (9)$$

where \mathbf{I} is the identity matrix, \dagger represents the complex conjugate transpose, P_v is the transmission power of node v , $E[\cdot]$ is the expectation function, and

$$\mathbf{E}[\mathbf{Z} \mathbf{Z}^\dagger] = \begin{bmatrix} \sigma_{r_0}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{r_1}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{r_{n-1}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{d_i}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{d_i}^2 \end{bmatrix}.$$

Expanding (9) gives us the mutual information as:

$$I_{CC}(s_i, \mathcal{R}, d_i) = \log_2 \left[1 + \frac{P_{s_i} |h_{s_i d_i}|^2}{\sigma_{d_i}^2} + \frac{P_{s_i} (\sum_{r_j \in \mathcal{R}} h_{r_j d_i} \alpha_{r_j} h_{s_i r_j})^2}{\sigma_{d_i}^2 + \sum_{r_j \in \mathcal{R}} |h_{r_j d_i}|^2 \alpha_{r_j}^2 \sigma_{r_j}^2} \right],$$

which can be rewritten as:

$$I_{CC}(s_i, \mathcal{R}, d_i) = \log_2 \left[1 + \text{SNR}_{s_i d_i} + \frac{\left(\sum_{r_j \in \mathcal{R}} \sqrt{\frac{\text{SNR}_{r_j d_i} \text{SNR}_{s_i r_j}}{1 + \text{SNR}_{s_i r_j}}} \right)^2}{1 + \sum_{r_j \in \mathcal{R}} \left(\frac{\text{SNR}_{r_j d_i}}{1 + \text{SNR}_{s_i r_j}} \right)} \right]. \quad (10)$$

Eq. (10) shows a complex relationship between a session's mutual information and the impact of each relay node. One observation that we can make now is that there is no clear conclusion (increase or decrease) on how the relay nodes affect the mutual information. It is entirely possible that adding more relay nodes in some setting will decrease the mutual information. This is because that when a relay node amplifies and forwards the received signal, it also amplifies and forwards the noise received along with the signal. As a result, if the background noise at a particular relay is very high, it can reduce the mutual information.

The achievable rate for session (s_i, d_i) can be written as:

$$\begin{aligned} C_{CC}(s_i, \mathcal{R}, d_i) &= \frac{W}{2} I_{CC}(s_i, \mathcal{R}, d_i) \\ &= \frac{W}{2} \log_2 \left[1 + \text{SNR}_{s_i d_i} + \frac{\left(\sum_{r_j \in \mathcal{R}} \sqrt{\frac{\text{SNR}_{r_j d_i} \text{SNR}_{s_i r_j}}{1 + \text{SNR}_{s_i r_j}}} \right)^2}{1 + \sum_{r_j \in \mathcal{R}} \left(\frac{\text{SNR}_{r_j d_i}}{1 + \text{SNR}_{s_i r_j}} \right)} \right]. \end{aligned} \quad (11)$$

In the special case when there is only one relay node in \mathcal{R} , the achievable rate in (11) reduces to (1).

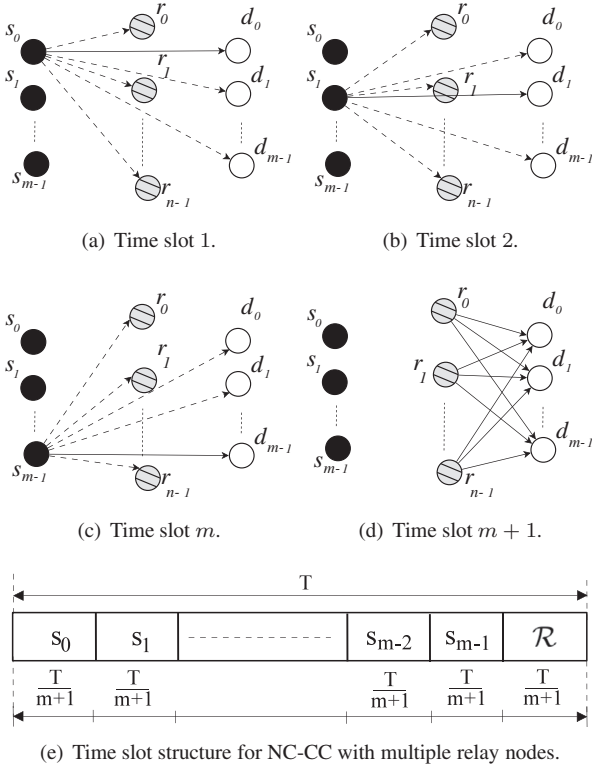


Fig. 4. Data transfer under multi-relay NC-CC in $m + 1$ time-slots.

3.2. Multiple sessions and multiple relays

We now extend our results in Section 3.1 to the general model where there are m sessions (denoted as $\mathcal{S} = \{s_0, s_1, \dots, s_{m-1}\}$) sharing n relay nodes with NC-CC.

The working of this general multi-session multi-relay NC-CC model is illustrated in Fig. 4. In this figure, the source node of each session transmits in the first m time slots, and each transmission is received by the destination nodes as well as the relay nodes. After the m transmissions, all the relay nodes will apply NC to combine the signals received during the m time slots. Next, Fig. 4d shows that the combined signal is amplified and simultaneously transmitted by all the relay nodes in the $(m + 1)$ th time slot.¹ The coherent signal reception at destination nodes can be facilitated by appropriate transmitter synchronization [16]. Fig. 4e shows the time-slot structure for this general multi-session multi-relay model. Every session is allotted a time slot duration of t . Thus, the time slot duration available to each source node is $\frac{T}{m+1} = \frac{|S|t}{|S|+1}$. Therefore, the achievable rate for a session (s_i, d_i) can be written as:

$$C_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i) = W \cdot \frac{\left(\frac{|S|t}{|S|+1}\right)}{|S|t} I_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i) \\ = \frac{W}{|S|+1} I_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i), \quad (12)$$

where $I_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i)$ is the mutual information between the source s_i and its destination node d_i .

Now we derive the mutual information $I_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i)$. First, the signal received at the destination node d_i in the $(m + 1)$ -th

time slot can be written as:

$$y_{\mathcal{S}\mathcal{R}d_i} = \sum_{r_j \in \mathcal{R}} \left(\alpha_{r_j} h_{r_j d_i} \sum_{s_k \in \mathcal{S}} y_{s_k r_j} \right) + z_{d_i} \\ = \sum_{r_j \in \mathcal{R}} \left(\alpha_{r_j} h_{r_j d_i} \sum_{s_k \in \mathcal{S}} (h_{s_k r_j} x_k + z_{r_j}) \right) + z_{d_i}, \quad (13)$$

where the value of amplification factor at r_j is:

$$\alpha_{r_j}^2 = \frac{P_{r_j}}{\sum_{s_k \in \mathcal{S}} (\sigma_{r_j}^2 + P_{s_k} |h_{s_k r_j}|^2)}.$$

A destination node d_i has to extract signal for x_i from the combined signal in (13). To extract the desired signal, a destination node d_i can subtract the signals received from other source nodes in the initial m time slots from this combined signal. The signals received by destination node d_i from other source nodes during the first m time slots can be written as:

$$y_{s_k d_i} = h_{s_k d_i} x_k + z_{d_i}, \quad (s_k \in \mathcal{S}, s_k \neq s_i). \quad (14)$$

To remove the signal x_k from the combined signal in (13), destination node d_i can multiply the overheard signal for x_k in (14) by $\frac{\sum_{r_j \in \mathcal{R}} \alpha_{r_j} h_{r_j d_i} h_{s_k r_j}}{h_{s_k d_i}}$, and then subtract this product from (13). When the signals for all x_k 's are removed from (13), the copy of the desired signal extracted by d_i can be written as:

$$\hat{y}_{\mathcal{S}\mathcal{R}d_i} = x_i \sum_{r_j \in \mathcal{R}} h_{s_i r_j} \alpha_{r_j} h_{r_j d_i} + \sum_{s_k \in \mathcal{S}} \sum_{r_j \in \mathcal{R}} \alpha_{r_j} h_{r_j d_i} z_{r_j} \\ + z_{d_i} - \sum_{s_k \in \mathcal{S}} \frac{\sum_{r_j \in \mathcal{R}} h_{s_k r_j} \alpha_{r_j} h_{r_j d_i}}{h_{s_k d_i}} h_{s_k d_i} \\ = \sum_{r_j \in \mathcal{R}} \alpha_{r_j} h_{r_j d_i} y_{s_i r_j} + z_{d_i} + \sum_{s_k \in \mathcal{S}} \sum_{r_j \in \mathcal{R}} \alpha_{r_j} h_{r_j d_i} z_{r_j} \\ - \sum_{s_k \in \mathcal{S}} \frac{\sum_{r_j \in \mathcal{R}} h_{s_k r_j} \alpha_{r_j} h_{r_j d_i}}{h_{s_k d_i}} z_{d_i}. \quad (15)$$

Due to the use of ANC and multiple relay nodes, we find in (15) that in addition to z_{d_i} , there are some new noise terms in the signal extracted by d_i . This new noise is called the ANC noise, and is written as:

$$z_{d_i}^{\text{ANC}} = z_{d_i} + \sum_{s_k \in \mathcal{S}} \sum_{r_j \in \mathcal{R}} \alpha_{r_j} h_{r_j d_i} z_{r_j} - \sum_{s_k \in \mathcal{S}} \frac{\sum_{r_j \in \mathcal{R}} h_{s_k r_j} \alpha_{r_j} h_{r_j d_i}}{h_{s_k d_i}} z_{d_i}.$$

We can now write the variance of ANC noise as:

$$\sigma_{z_{d_i}^{\text{ANC}}}^2 = \sigma_{d_i}^2 + (|\mathcal{S}| - 1) \sum_{r_j \in \mathcal{R}} (\alpha_{r_j} h_{r_j d_i})^2 \sigma_{r_j}^2 \\ \times \sum_{s_k \in \mathcal{S}} \sum_{r_j \in \mathcal{R}} \left(\frac{h_{s_k r_j} \alpha_{r_j} h_{r_j d_i}}{h_{s_k d_i}} \right)^2 \sigma_{d_i}^2. \quad (16)$$

From (16), we can make two important observations: (i) the variance in ANC noise increases monotonically as the group size of sessions sharing the same set of relays increase; and (ii) the variance in ANC noise increases monotonically with the size of the set of relays.

Similar to Section 3.1, we can now write (6) and (15) in a compact matrix form as follows:

$$\hat{\mathbf{Y}} = \mathbf{H} \mathbf{x}_i + \mathbf{B} \hat{\mathbf{Z}},$$

¹ Note that the scheme in [29] needs $(m + n)$ time slots and thus yields a smaller rate.

where

$$\hat{\mathbf{Y}} = \begin{bmatrix} y_{s_i d_i} \\ \hat{y}_{s_i \mathcal{R} d_i} \end{bmatrix} \text{ and } \hat{\mathbf{Z}} = \begin{bmatrix} z_{r_0} \\ z_{r_1} \\ \vdots \\ z_{r_{n-1}} \\ z_{d_i} \\ z_{d_i}^{\text{ANC}} \end{bmatrix}.$$

Similar to the one-session multi-relay case in Section 3.1, we can model the above channel from $s_i \rightarrow d_i$ and from $s_i \rightarrow \mathcal{R} \rightarrow d_i$ as a one-input two-output complex Gaussian channel. The mutual information between s_i and d_i can be written as:

$$I_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i) = \log \det \{ \mathbf{I} + (P_{s_i} \mathbf{H} \mathbf{H}^\dagger) (\mathbf{B} \mathbf{E} [\hat{\mathbf{Z}} \hat{\mathbf{Z}}^\dagger] \mathbf{B}^\dagger)^{-1} \}, \quad (17)$$

where

$$\mathbf{E}[\hat{\mathbf{Z}} \hat{\mathbf{Z}}^\dagger] = \begin{bmatrix} \sigma_{r_0}^2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \sigma_{r_1}^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{r_{n-1}}^2 & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{d_i}^2 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \sigma_{z_{d_i}^{\text{ANC}}}^2 \end{bmatrix}.$$

Expanding (17) gives us the value of mutual information between s_i and d_i as follows:

$$I_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i) = \log_2 \left[1 + \frac{P_{s_i} |h_{s_i d_i}|^2}{\sigma_{d_i}^2} + \frac{P_{s_i} \left(\sum_{r_j \in \mathcal{R}} h_{s_i r_j} \alpha_{r_j} h_{r_j d_i} \right)^2}{\sigma_{z_{d_i}^{\text{ANC}}}^2 + \sum_{r_j \in \mathcal{R}} h_{r_j d_i}^2 \alpha_{r_j}^2 \sigma_{r_j}^2} \right],$$

which can be rewritten as:

$$I_{\text{NC-CC}}(s_i, \mathcal{S}, \mathcal{R}, d_i) = \log_2 \left[1 + \text{SNR}_{s_i d_i} + \frac{\left(\sum_{r_j \in \mathcal{R}} \sqrt{\frac{\text{SNR}_{r_j d_i} \text{SNR}_{s_i r_j}}{|\mathcal{S}| + \sum_{s_k \in \mathcal{S}} \text{SNR}_{s_k r_j}}} \right)^2}{\frac{\sigma_{z_{d_i}^{\text{ANC}}}^2}{\sigma_{d_i}^2} + \sum_{r_j \in \mathcal{R}} \left(\frac{\text{SNR}_{r_j d_i}}{|\mathcal{S}| + \sum_{s_k \in \mathcal{S}} \text{SNR}_{s_k r_j}} \right)} \right]. \quad (18)$$

From (18), we can see that the value of mutual information depends directly on the individual relay nodes in \mathcal{R} and the value of ANC noise. We can make two important observations by taking a close look at (18). First, a given session (s_i, d_i) performing NC-CC with a group of relays (r_j 's) would prefer these relays to stay as close as possible to s_i . Intuitively, this is true because the signal that r_j 's will receive from s_i will have smaller noise component (or larger SNR) due to proximity to s_i . A mathematical explanation of this is that the term $\text{SNR}_{s_i r_j}$ is in the denominator of (18). Second, s_i would also prefer to have other s_k 's in the session group to be far away from the relay nodes. This is because when the distance between the other source nodes (i.e., other s_k 's) and the relay nodes increases, the value of channel gains between other s_k 's and the relays become small, which will help reduce the ANC noise value at destination d_i . This can be observed in (16) where the value of ANC noise decreases monotonically with the decrease in the value of $h_{s_k r_j}$.

3.3. An example

We now apply the achievable rate Eqs. (12) and (18) in a small network to study how the data rates of each session change when different set of relay nodes are employed.

Consider the network topology in Fig. 5 where there are two sessions (s_0, d_0) and (s_1, d_1) . There are four relay nodes that can

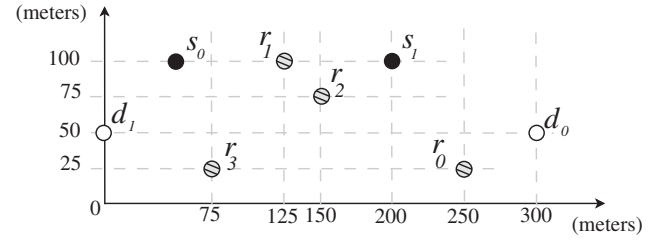


Fig. 5. Two-session four-relay network.

Table 2

Data rates under NC-CC with different sets of relay nodes.

\mathcal{R}	Data Rate (Mbps)		Total (Mbps)
	(s_0, d_0)	(s_1, d_1)	
$\mathcal{R}_0 = \emptyset$	16.70	25.83	42.53
$\mathcal{R}_1 = \{r_0\}$	13.65	20.01	33.66
$\mathcal{R}_2 = \{r_1\}$	18.48	27.44	45.92
$\mathcal{R}_3 = \{r_2\}$	13.81	28.80	42.61
$\mathcal{R}_4 = \{r_3\}$	17.50	23.14	40.64
$\mathcal{R}_5 = \{r_0, r_1\}$	15.45	31.78	47.23
$\mathcal{R}_6 = \{r_0, r_2\}$	14.63	33.09	47.72
$\mathcal{R}_7 = \{r_0, r_3\}$	15.22	26.07	41.29
$\mathcal{R}_8 = \{r_1, r_2\}$	22.30	37.82	60.12
$\mathcal{R}_9 = \{r_1, r_3\}$	26.24	30.03	56.27
$\mathcal{R}_{10} = \{r_2, r_3\}$	21.57	30.63	52.19
$\mathcal{R}_{11} = \{r_0, r_1, r_2\}$	16.52	40.53	57.04
$\mathcal{R}_{12} = \{r_0, r_1, r_3\}$	17.15	32.39	49.54
$\mathcal{R}_{13} = \{r_0, r_2, r_3\}$	16.28	32.97	49.25
$\mathcal{R}_{14} = \{r_1, r_2, r_3\}$	28.49	35.96	64.45
$\mathcal{R}_{15} = \{r_0, r_1, r_2, r_3\}$	18.22	37.77	55.99

be employed, namely, r_0, r_1, r_2 and r_3 . We assume that the channel bandwidth in the network is $W = 20$ MHz, the white Gaussian noise at all the nodes has a variance of $10^{-10}W$, and every node transmits at a power of 1 W. We assume the channel gain between two nodes u and v is $\|u - v\|^{-4}$, where $\|u - v\|$ is the distance (in meters) between u and v , the path loss index is 4.

We calculate the achievable rate of each session under NC-CC when different set of relay nodes are used. Since there are four relay nodes, there are 16 possibilities (see first column in Table 2), ranging from the trivial case of no relay node is used to the extreme case when all four relay nodes are used. Table 2 lists the data rates for each session under each of these 16 cases. When $\mathcal{R} = \emptyset$, Eq. (2) (for direct transmission) is used and the effective bandwidth available to each session is $\frac{W}{2} = 10$ MHz. The last column in Table 2 shows the sum of the data rates of both sessions.

For the 15 sets of relay nodes (excluding direct transmission), we observe that not every set is beneficial to both sessions, compared to the case of direct transmission. For session (s_0, d_0) , the sets of $\mathcal{R}_1, \mathcal{R}_3, \mathcal{R}_5, \mathcal{R}_6, \mathcal{R}_7, \mathcal{R}_{11}$, and \mathcal{R}_{13} are not beneficial, whereas for session (s_1, d_1) , the sets of \mathcal{R}_1 and \mathcal{R}_4 are not beneficial. Some sets of relay nodes are beneficial to one session but not to the other session, e.g., relay groups $\mathcal{R}_3, \mathcal{R}_5, \mathcal{R}_6, \mathcal{R}_7, \mathcal{R}_{11}$, and \mathcal{R}_{13} are beneficial to (s_1, d_1) but not to (s_0, d_0) . Also, the most beneficial set of relay nodes for (s_0, d_0) is \mathcal{R}_{14} , and the most beneficial set of relay nodes for (s_1, d_1) is \mathcal{R}_{11} .

Furthermore, as per our discussion at the end of Section 3, we can observe that \mathcal{R}_4 is beneficial to session (s_0, d_0) but not to session (s_1, d_1) due to its proximity to source s_0 . Similarly, \mathcal{R}_3 is beneficial to (s_1, d_1) but not to (s_0, d_0) due to its proximity to s_1 compared to s_0 . For some other relay groups, the conclusion may not be made simply by observing the location of individual nodes. This is where our derived equations can help in determining whether a session group should be matched with a relay group or not. It is

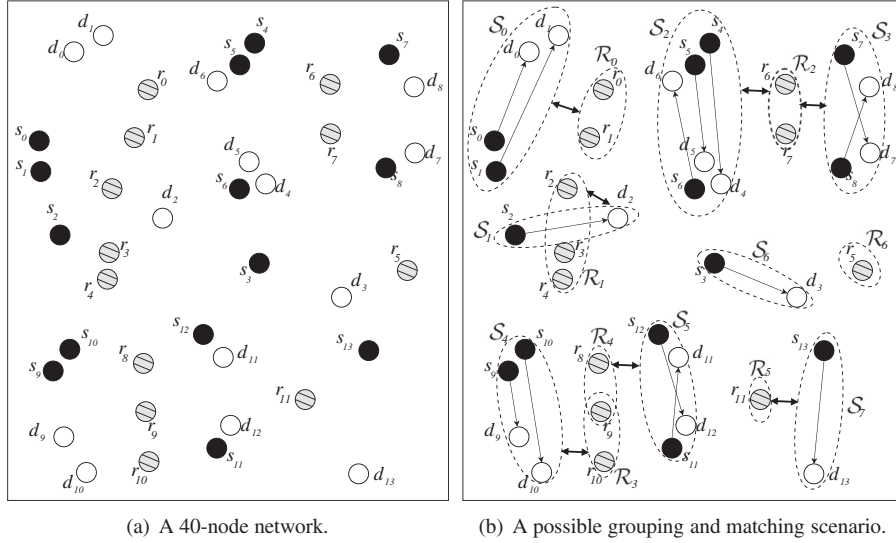


Fig. 6. An example illustrating our session/relay grouping and matching problem.

important to realize that without deriving the mathematical equations for achievable data rate and mutual information, it is impossible to construct a table such as the one shown in Table 2. The formulas we developed can thus help us to identify the optimal set of relay nodes for some particular objective. We will next study an optimization problem that will use the derived equations to achieve a certain objective.

4. Problem description

Table 1 lists all notation in this paper. Consider a network (e.g., Fig. 6a) where there is a set of sessions $S = \{(s_0, d_0), (s_1, d_1), \dots, (s_{m-1}, d_{m-1})\}$ and a set of relay nodes $\mathcal{R} = \{r_0, r_1, \dots, r_{n-1}\}$. For each session (s_i, d_i) , the source node s_i always have data to transmit to the destination node d_i .² Assume all the nodes are in the same interference (collision) domain. Therefore, similar to Fig. 4e, a time frame of length T needs to be divided among the sessions to coordinate transmissions. Given the availability of relay nodes, NC-CC may be used. Our goal is to exploit the potential of NC-CC and set up a transmission schedule so that some network-wide objective is maximized.

In this network setting, a number of questions arises naturally.

- First and foremost, from each session's perspective, what set of relay nodes should it employ to increase its achievable rate?
- Second, from each relay node's perspective, what set of sessions should it support (in the context of NC-CC)?
- Third, should we partition the set of sessions and relay nodes into different groups? And if so, how to group these sessions and relay nodes, and how to match them to maximize our objective?
- Finally, how should the time slots be structured in a time frame so as to coordinate the transmissions of all the sessions?

Regarding the first question, one can quickly deduce, by a simple numerical analysis of (18), that blind employment of all relay nodes in the network may not maximize a session's achievable rate. This is because background noise, introduced in the received signals at certain relay nodes, could be high. Once such noisy signal is amplified, transmitted, and aggregated with signals from

other relay nodes, it will lead to large noise in the received signal at the destination node, thereby reducing the session's achievable rate. Another issue is that the variance of NC noise increases monotonically as the size of the relay group. Therefore, in the interest of each session, it is important to select an optimal subset of relay nodes to maximize its achievable rate.

For the second question, by observing (16), one can easily find that the variance of NC noise increases monotonically as the number of sessions. Since the achievable rate decreases as NC noise variance increases, we conclude that loading a relay node with a large number of sessions will not maximize our objective. Therefore, from a relay node's perspective, it is important to select an optimal subset of sessions.

Based on the above discussion, it is easy to answer the third question. Clearly, we need to partition the set of sessions and relay nodes into different groups. Note that there could be some overlap among the sets of relay nodes, i.e., a relay node may be in multiple groups. However, a session can only appear in one group. As we shall show, grouping of sessions and relay nodes is not an easy task, neither is the problem of matching them to maximize our objective.

For the last question, once the optimal session/relay grouping and matching problem is solved, the time slot structure will also be determined.

We now use an example to illustrate the scope of our problem.

4.1. An example

Consider again the network in Fig. 6a. There are 14 sessions and 12 relay nodes. Fig. 6b shows one possible grouping and matching solution (although may not be optimal) for the network, and Fig. 7 shows the time slot structure for this solution. The following are some details in this example solution.

Groups of sessions. There are eight session groups: $S_0 = \{(s_0, d_0), (s_1, d_1)\}$, $S_1 = \{(s_2, d_2)\}$, $S_2 = \{(s_4, d_4), (s_5, d_5), (s_6, d_6)\}$, $S_3 = \{(s_7, d_7), (s_8, d_8)\}$, $S_4 = \{(s_9, d_9), (s_{10}, d_{10})\}$, $S_5 = \{(s_{11}, d_{11}), (s_{12}, d_{12})\}$, $S_6 = \{(s_3, d_3)\}$, $S_7 = \{(s_{13}, d_3)\}$. Note that each session can belong to only one group.

Groups of relays. There are seven relay groups: $\mathcal{R}_0 = \{r_0, r_1\}$, $\mathcal{R}_1 = \{r_2, r_3, r_4\}$, $\mathcal{R}_2 = \{r_6, r_7\}$, $\mathcal{R}_3 = \{r_9, r_{10}\}$, $\mathcal{R}_4 = \{r_8, r_9\}$, $\mathcal{R}_5 = \{r_{11}\}$, $\mathcal{R}_6 = \{r_5\}$. Note that relay node r_9 shows up both in \mathcal{R}_3 and \mathcal{R}_4 .

² We assume that each node can only serve one distinct role of source, destination, or relay. In the case when a node is serving multiple roles, we can logically divide this node into multiple nodes and consider each separately.

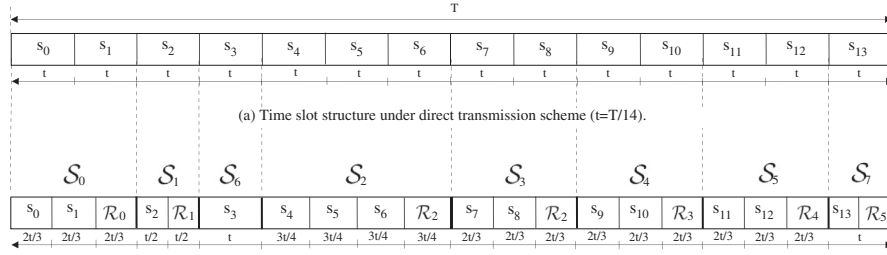


Fig. 7. Time slot structure for the 40 node network under different schemes.

Matching session and relay groups. In this example, s_0 is matched to \mathcal{R}_0 . Similarly, s_1 is matched to \mathcal{R}_1 . But, s_2 and s_3 are both matched to \mathcal{R}_2 , which is allowed. s_4, s_5 , and s_7 are matched to $\mathcal{R}_3, \mathcal{R}_4$ and \mathcal{R}_5 , respectively. s_6 is not matched to any relay group while \mathcal{R}_6 is not used.

Time slot structure. Fig. 7b shows the time slot structure for this session/relay grouping and matching example. For comparison, Fig. 7a shows the time slot structure when NC-CC is not employed, i.e., under TDMA-based direct transmission. In Fig. 7b, since session s_3 does not use any relay node, its time slot length should be kept same as that in Fig. 7a to ensure fairness. On the other hand, for those session groups that use relay nodes, the time slot for each session should be shrunk accordingly to accommodate the extra time slot for the relay nodes. For example, since sessions s_0 and s_1 are using \mathcal{R}_0 , their time slot length of t (under direct transmission) is shrunk to $\frac{2}{3}t$ so that the three time slots taken by s_0, s_1 and \mathcal{R}_0 remain $2t$. In general, a set \mathcal{S}_k with $|\mathcal{S}_k|$ sessions will have total available time of $|\mathcal{S}_k|t$. When this session group uses a relay group \mathcal{R}_j for NC-CC, the time slot duration for every session in \mathcal{S}_k will be shrunk to $\frac{|\mathcal{S}_k|t}{|\mathcal{S}_k|+1}$.

Achievable rate. Based on the above discussion on time slot structure, the achievable rate for a session (s_i, d_i) is

$$\begin{aligned}
 C_{\text{NC-CC}}(s_i, \mathcal{S}_i^{\mathcal{R}_j}, \mathcal{R}_j, d_i) &= W \cdot \frac{\left(\frac{|\mathcal{S}_i^{\mathcal{R}_j}|t}{|\mathcal{S}_i^{\mathcal{R}_j}|+1} \right)}{|\mathcal{S}_i|t} \cdot I_{\text{NC-CC}}(s_i, \mathcal{S}_i^{\mathcal{R}_j}, \mathcal{R}_j, d_i) \\
 &= \frac{|\mathcal{S}_i^{\mathcal{R}_j}|}{|\mathcal{S}_i^{\mathcal{R}_j}|+1} \cdot \frac{W}{|\mathcal{S}_i|} \cdot I_{\text{NC-CC}}(s_i, \mathcal{S}_i^{\mathcal{R}_j}, \mathcal{R}_j, d_i) \\
 \mathcal{S}_i^{\mathcal{R}_j} &\subseteq \mathcal{S}_i, s_i \in \mathcal{S}_i^{\mathcal{R}_j},
 \end{aligned} \tag{19}$$

where $\mathcal{S}_i^{\mathcal{R}_j}$ denotes the session group (containing session (s_i, d_i)) that is matched to relay nodes in \mathcal{R}_j .

4.2. Problem complexity

Our goal in this paper is to perform optimal grouping of sessions and relay nodes, and matching these groups so that the sum of weighted session rates is maximized. For session grouping, the smaller the size of each group, the larger the mutual information, due to smaller NC noise. But on the other hand, comparing (12) with (19), we find that smaller session group size will also have smaller effective bandwidth (i.e., $\frac{|\mathcal{S}_i^{\mathcal{R}_j}|}{|\mathcal{S}_i^{\mathcal{R}_j}|+1} \frac{W}{|\mathcal{S}_i|} < \frac{W}{|\mathcal{S}_i|+1}$ because

$|\mathcal{S}_i|$ is greater than $|\mathcal{S}_i^{\mathcal{R}_j}|$). For grouping of relay nodes, there is even more flexibility, as any relay node may be part of multiple relay groups. Finally, the optimal matching problem is highly complex, due to the large design space of potential session groups and relay groups.

Theorem 1. The joint session/relay grouping and matching problem for NC-CC is NP-hard.

A sketch of proof is as follows. In [26], Sharma et al. considered a simpler grouping and relay node selection (GRS) problem, with the same objective of maximizing the weighted sum rate of the sessions in the network. There was no consideration of grouping of relay nodes. In other words, the size of each relay group was set to 1, which can be viewed as a special case of the problem in this paper. For the GRS problem, Sharma et al. used matching problems in hypergraphs to show that the GRS problem is NP-hard. Given that the GRS problem is a special case of our joint session/relay grouping and matching problem, we conclude that our problem is at least NP-hard.

5. G²M: an algorithm for session/relay grouping and matching

In this section, we present an algorithm that performs session grouping, relay grouping, and matching of session and relay groups. We abbreviate this algorithm as G²M, with the “2” referring that grouping operation is needed for both sessions and relay nodes. We present the baseline G²M algorithm in Sections 5.1–5.4. Further improvements of the baseline algorithm are discussed in Section 5.5.

5.1. Basic idea

The basic idea of G²M is to have each session initially matched independently to a group of relay nodes. Then through merging of sessions and modifications of relay node groups iteratively, we obtain a final solution. Fig. 8 shows the flow chart of the G²M algorithm.

In the initialization phase, we let each session (s_i, d_i), $i = 0, \dots, m-1$, form a group on its own, i.e., $\mathcal{S}_i = \{(s_i, d_i)\}$, $i = 0, \dots, m-1$. Then for each session group \mathcal{S}_i (which has only one session), we find a set of relay nodes for it, which we denote as \mathcal{R}_i , $i = 0, \dots, m-1$. The set of relay nodes is determined through an iterative process that begins by considering all the relay nodes in the set, and then removing some relay nodes from the set that are harmful for that particular session.

In the main program, during each iteration, we consider pairwise of session groups and see if merging the two will result in an improved objective function. Clearly, merging of two session groups also requires the merging of two groups of relay nodes. To increase the chance of successful merger of two session groups, modifications of relay nodes (in terms of removing some nodes) are allowed in the newly merged relay node groups. Such iteration terminates when we cannot find a pair of session groups to merge that can produce a greater objective value. At this point, G²M terminates.

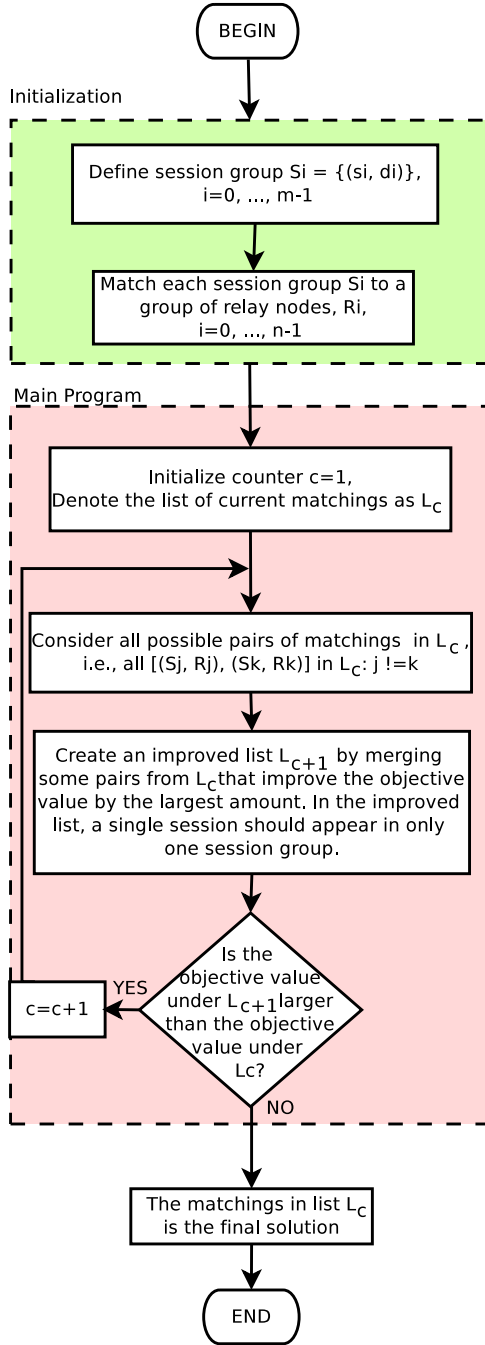


Fig. 8. A flowchart of G²M algorithm.

5.2. Algorithm details

Initialization. As discussed in Section 5.1, we start with each session group containing only one session, i.e. $S_i = \{(s_i, d_i)\}$, $i = 0, \dots, m-1$. For each session S_i , we will find a group of relay nodes \mathcal{R}_i , $i = 0, \dots, m-1$ for it so that the achievable rate of this session is maximized. Based on (18), when $S_i = \{(s_i, d_i)\}$, is matched to a relay group \mathcal{R}_i , its mutual information is

$$I_{\text{NC-CC}}(S_i, \{(s_i, d_i)\}, \mathcal{R}_i, d_i) = \log_2 \left(1 + \text{SNR}_{s_i d_i} + \frac{\left(\sum_{r_j \in \mathcal{R}_i} \sqrt{\frac{\text{SNR}_{r_j d_i} \text{SNR}_{s_i r_j}}{1 + \text{SNR}_{s_i r_j}}} \right)^2}{1 + \sum_{r_j \in \mathcal{R}_i} \left(\frac{\text{SNR}_{r_j d_i}}{1 + \text{SNR}_{s_i r_j}} \right)} \right). \quad (20)$$

From (20), the SNR-gain for (s_i, d_i) due to the group of relay node \mathcal{R}_i is

$$\text{SNR}_{\text{gain}}(S_i, \mathcal{R}_i) = \frac{\left(\sum_{r_j \in \mathcal{R}_i} \sqrt{\frac{\text{SNR}_{r_j d_i} \text{SNR}_{s_i r_j}}{1 + \text{SNR}_{s_i r_j}}} \right)^2}{1 + \sum_{r_j \in \mathcal{R}_i} \left(\frac{\text{SNR}_{r_j d_i}}{1 + \text{SNR}_{s_i r_j}} \right)}. \quad (21)$$

Now we need to find a group of relay nodes \mathcal{R}_i for each S_i that can maximize SNR_{gain} . The following theorem shows that this problem is also NP-hard.

Theorem 2. For a single session, the problem of finding an optimal group of relay nodes that maximizes the session's achievable rate is NP-hard.

We offer a sketch of proof here. In [26], Sharma et al. showed that the problem of having a single relay node to select an optimal group of sessions among a set of sessions is NP-hard. The proof technique there was based on matching problems in hypergraphs. The mathematical nature of that problem is exactly the same as this one, and thus the proof here can follow the same token.

We now present a heuristic algorithm to construct an initial matching.

- For $S_i = (s_i, d_i)$, we start with the group \mathcal{R}_i including all the relay nodes in the network.
- To maximize SNR_{gain} , we identify and remove certain relay nodes (one at a time) from \mathcal{R}_i .
 - The first candidate for possible removal is the relay node with the worst channel condition between s_i and itself, i.e., the relay node r_j with the smallest value of $\text{SNR}_{s_i r_j}$. This node is likely to introduce the largest noise component. From (20), we can see that this relay node is also likely to contribute the largest amount to the denominator.
 - Remove this relay node, say r_j with the smallest $\text{SNR}_{s_i r_j}$. If session (s_i, d_i) 's mutual information increase, this removal is permanent; otherwise, r_j is added back to \mathcal{R}_i .³
 - Repeat the above process for the relay node with the second smallest value of $\text{SNR}_{s_i r_j}$ and so forth.
- During the above iteration for S_i , some relay nodes may be removed from \mathcal{R}_i . As a result, we should go through another iteration of checking and removing the relay nodes from the current \mathcal{R}_i . This is due to the nonlinear nature of (21). Note that in the above iteration, when we checked the current relay nodes in \mathcal{R}_i for removal, the relay nodes in \mathcal{R}_i at that time were different from the current \mathcal{R}_i . Thus, it may now be possible to further remove some of the current relay nodes and improve \mathcal{R}_i .
- We stop the process of removing the relay nodes from \mathcal{R}_i until during an iteration, none of the relay nodes are removed from \mathcal{R}_i . The current relay nodes in \mathcal{R}_i constitutes the initial group of relay nodes that is matched to $S_i = (s_i, d_i)$.
- As a last step, we want to ensure that the achievable rate of session (s_i, d_i) is no less than that under direct transmission. If yes, we are done; otherwise, we set $\mathcal{R}_i = \emptyset$, indicating that initially no relay node will be matched to this session.

Fig. 9 gives the pseudocode for the initial matching between each session and its relay nodes.

Main program. After initialization, we now have an initial list (say \mathcal{L}_1) of $m = |S|$ matchings with every session group (containing a single session) matched to a group of relay nodes. Note that NC is not yet employed and the goal of the main program is to merge session groups (two at a time) so that NC can be fully exploited to

³ This is because that, from (21), not only the values of $\text{SNR}_{s_i r_j}$, but also the values of $\text{SNR}_{r_j d_i}$ and the SNR values of the other relay nodes are affecting the value of SNR_{gain} .

Initialize()	
1.	For every $(s_i, d_i) \in \mathcal{S}$
2.	Set $\mathcal{R}_{beg} = \emptyset, \hat{\mathcal{R}} = \mathcal{R}$
3.	While $\mathcal{R}_{beg} \neq \hat{\mathcal{R}}$
4.	$\mathcal{R}_{beg} = \hat{\mathcal{R}}$
5.	Consider all $r_j \in \hat{\mathcal{R}}$ in increasing order of $\text{SNR}_{s_i r_j}$ values
6.	if $(C_{\text{NC-CC}}^{w_i}(s_i, \{s_i\}, \hat{\mathcal{R}}, d_i) > C_{\text{NC-CC}}^{w_i}(s_i, \{s_i\}, \hat{\mathcal{R}} - r_j, d_i)$
7.	Then, $\hat{\mathcal{R}} = \hat{\mathcal{R}} - r_j$
8.	If $(C_{\text{NC-CC}}^{w_i}(s_i, \{s_i\}, \hat{\mathcal{R}}, d_i) < (C_{\text{D}}^{w_i}(s_i, d_i)$, then set $\hat{\mathcal{R}} = \emptyset$.
9.	Set $\hat{\mathcal{R}}$ as the relay group matched with $\{(s_i, d_i)\}$.

Fig. 9. Pseudocode of initialization phase.

increase the objective of our optimization problem (i.e., weighted sum rate of all sessions).

In the first iteration, we go through the initial list \mathcal{L}_1 that has m entries of (S_i, \mathcal{R}_i) matchings. We consider all possible pairs of entries (S_i, \mathcal{R}_i) and (S_j, \mathcal{R}_j) , $S_i \neq S_j$, for possible merger. There are $\frac{m(m-1)}{2}$ possibilities.

Denote $\mathcal{L}_{\text{temp}}$ a temporary working list to store our intermediate matching results. For every matching pair of entries $[(S_i, \mathcal{R}_i), (S_j, \mathcal{R}_j)]$ in \mathcal{L}_1 , we perform the following steps.

- Suppose two session groups S_i and S_j were merged together into one session group $S_i \cup S_j$. Then the two corresponding relay node groups \mathcal{R}_i and \mathcal{R}_j are also merged together into one relay node group $\mathcal{R}_i \cup \mathcal{R}_j$. Now we have a new session group $S_i \cup S_j$ matched to a new relay group $\mathcal{R}_i \cup \mathcal{R}_j$.
- Given that \mathcal{R}_i and \mathcal{R}_j are likely containing different set of nodes, some of which may benefit sessions in one group but not the other. To ensure that every relay node in $\mathcal{R}_i \cup \mathcal{R}_j$ will benefit the new session group $S_i \cup S_j$, we examine each non-overlapping relay node in $\mathcal{R}_i \cup \mathcal{R}_j$ (i.e., the relay nodes that are not part of both \mathcal{R}_i and \mathcal{R}_j) one at a time and remove any relay, say r_k if its presence in $\mathcal{R}_i \cup \mathcal{R}_j$ is harmful to the objective function for the new session group $S_i \cup S_j$. After this process, we have an updated relay group, which we denote as $(\mathcal{R}_i \cup \mathcal{R}_j)^*$.
- To determine whether or not the proposed new matching $(S_i \cup S_j, (\mathcal{R}_i \cup \mathcal{R}_j)^*)$ should be stored in $\mathcal{L}_{\text{temp}}$, we compare whether or not there is any improvement in the objective function, i.e., whether or not $\sum_{s_k \in S_i \cup S_j} w_k C_{\text{NC-CC}}(s_k, S_i \cup S_j, (\mathcal{R}_i \cup \mathcal{R}_j)^*, d_k) > (\sum_{s_k \in S_i} w_k C_{\text{NC-CC}}(s_k, S_i, \mathcal{R}_i, d_k) + \sum_{s_k \in S_j} w_k C_{\text{NC-CC}}(s_k, S_j, \mathcal{R}_j, d_k))$?
 - If there is an increase in objective, then we store the new matching $(S_i \cup S_j, (\mathcal{R}_i \cup \mathcal{R}_j)^*)$ in $\mathcal{L}_{\text{temp}}$. Also, we calculate the net increase in the objective value due to this merger, which we call *temporary gain*.
 - If the objective value decreases or remains same, then there is no benefit in merging S_i and S_j . Therefore, we declare this proposed merger a failure. If (S_i, \mathcal{R}_i) and (S_j, \mathcal{R}_j) have not been stored in $\mathcal{L}_{\text{temp}}$, we will store both as two entries in $\mathcal{L}_{\text{temp}}$ and associate each with a zero temporary gain.

We now have a list $\mathcal{L}_{\text{temp}}$ containing several beneficial matchings and some matchings with gain zero. Note that a session group S_i may be part of multiple matchings in list $\mathcal{L}_{\text{temp}}$. We now want to create a list \mathcal{L}_2 where any session group S_i will only appear in exactly one matching. This is equivalent to have each session appear only once in some session group in \mathcal{L}_2 . To accomplish this, we consider entries in $\mathcal{L}_{\text{temp}}$ in decreasing value of temporary gain. For any such entry under consideration, we do the following.

- If none of the sessions in this session group appears in any session group of \mathcal{L}_2 , this entry of matching (session group and relay node group) is then saved in \mathcal{L}_2 . This entry is also removed from $\mathcal{L}_{\text{temp}}$.

Table 3
Initial matching list \mathcal{L}_1 .

Session group	Relay node group
$S_0 = \{(s_0, d_0)\}$	$\mathcal{R}_0 = \{r_0, r_1\}$
$S_1 = \{(s_1, d_1)\}$	$\mathcal{R}_1 = \{r_1, r_2, r_3\}$
$S_2 = \{(s_2, d_2)\}$	$\mathcal{R}_2 = \{r_3\}$
$S_3 = \{(s_3, d_3)\}$	$\mathcal{R}_3 = \{\emptyset\}$
$S_4 = \{(s_4, d_4)\}$	$\mathcal{R}_4 = \{r_1, r_4\}$

Table 4
All possible matching pairs from \mathcal{L}_1 .

	Pair under consideration	Proposed merger	CIO
1	$\{(S_0, \mathcal{R}_0), (S_1, \mathcal{R}_1)\}$	$\{(s_0, d_0), (s_1, d_1)\}, \{r_0, r_1, r_2\}$	15
2	$\{(S_0, \mathcal{R}_0), (S_2, \mathcal{R}_2)\}$	$\{(s_0, d_0), (s_2, d_2)\}, \{r_1, r_3\}$	10
3*	$\{(S_0, \mathcal{R}_0), (S_3, \mathcal{R}_3)\}$	$\{(s_0, d_0), (s_3, d_3)\}, \{\emptyset\}$	-5
4	$\{(S_0, \mathcal{R}_0), (S_4, \mathcal{R}_4)\}$	$\{(s_0, d_0), (s_4, d_4)\}, \{r_0, r_1, r_4\}$	4
5	$\{(S_1, \mathcal{R}_1), (S_2, \mathcal{R}_2)\}$	$\{(s_1, d_1), (s_2, d_2)\}, \{r_1, r_3\}$	7
6*	$\{(S_1, \mathcal{R}_1), (S_3, \mathcal{R}_3)\}$	$\{(s_1, d_1), (s_3, d_3)\}, \{r_1\}$	-10
7*	$\{(S_1, \mathcal{R}_1), (S_4, \mathcal{R}_4)\}$	$\{(s_1, d_1), (s_4, d_4)\}, \{r_2, r_4\}$	-9
8	$\{(S_2, \mathcal{R}_2), (S_3, \mathcal{R}_3)\}$	$\{(s_2, d_2), (s_3, d_3)\}, \{r_3\}$	1
9	$\{(S_2, \mathcal{R}_2), (S_4, \mathcal{R}_4)\}$	$\{(s_2, d_2), (s_4, d_4)\}, \{r_1\}$	6
10	$\{(S_3, \mathcal{R}_3), (S_4, \mathcal{R}_4)\}$	$\{(s_3, d_3), (s_4, d_4)\}, \{r_4\}$	3

- If all sessions in this session group already appear in some session groups in \mathcal{L}_2 , this entry is not saved in \mathcal{L}_2 . Further, this entry is also removed from $\mathcal{L}_{\text{temp}}$.
- If some, but not all, sessions of this session group appear in some session groups in \mathcal{L}_2 , then we will recover the session group containing the remaining sessions (i.e., those not shown up in \mathcal{L}_2) and its matching relay group from \mathcal{L}_1 . This recovered matching entry will carry a temporary gain of zero and will replace the one in $\mathcal{L}_{\text{temp}}$.

The above process continues until $\mathcal{L}_{\text{temp}}$ is empty. At this point, each session should appear only once in some session group in \mathcal{L}_2 . This completes the first iteration of our main program.

The future iterations of the main program are similar to the first iteration. The program terminates when no further mergers are possible, i.e., the temporary gain is zero for all entries in $\mathcal{L}_{\text{temp}}$. Then the matching created in the previous iteration is our final solution. Fig. 10 shows the pseudocode of the main program.

5.3. An example

We now present an example showing the working of the G²M algorithm. Consider a network with five sessions $\{(s_0, d_0), \dots, (s_4, d_4)\}$ and five relay nodes $\{r_0, \dots, r_4\}$. Table 3 shows the initial matching \mathcal{L}_1 created during the initialization phase. Note that a session may use several relay nodes, and a relay node may be matched to multiple sessions (e.g., r_1 and r_3).

Based on \mathcal{L}_1 , we construct a list $\mathcal{L}_{\text{temp}}$ to facilitate the creation of list \mathcal{L}_2 for the next iteration. Table 4 shows all possible matching pairs from \mathcal{L}_1 (in second column), the proposed mergers (in third column), and the corresponding changes in the objective value (in the last column). $\mathcal{L}_{\text{temp}}$ is constructed based on Table 4 as follows:

- Entries 1 and 2 in Table 4 offer an increase in objective value. So, we copy these two entries to $\mathcal{L}_{\text{temp}}$ in Table 5.
- Entry 3 in Table 4 leads to a decrease in the objective value. So, this proposed merger is not desirable and will not show up in $\mathcal{L}_{\text{temp}}$. Instead, we will include the matchings prior to the proposed merger to $\mathcal{L}_{\text{temp}}$ in Table 5. The corresponding gains in the objective value for these two entries are zero.
- Similarly, we will include Entries 4, 5, 8, 9, and 10 from Table 4 to $\mathcal{L}_{\text{temp}}$ in Table 5, as they all offer an increase in the objective value.

Main()	
1.	Set $k=1, \mathcal{L}_0 = \emptyset$
2.	Initialize()
3.	For all $s_i \in \mathcal{S}$,
4.	Add (s_i, d_i) to \mathcal{S}_i
5.	Set \mathcal{R}_i as the relay group for \mathcal{S}_i constructed during the initialization phase
6.	Add the matching $(\mathcal{S}_i, \mathcal{R}_i)$ to \mathcal{L}_k
7.	Set $\text{merge_count}=0, \mathcal{L}_{\text{temp}} = \emptyset$
8.	For all pairs of matchings $[(\mathcal{S}_i, \mathcal{R}_i), (\mathcal{S}_j, \mathcal{R}_j)]$ in \mathcal{L}_k , such that either $(\mathcal{S}_i, \mathcal{R}_i) \notin \mathcal{L}_{k-1}$ or $(\mathcal{S}_j, \mathcal{R}_j) \notin \mathcal{L}_{k-1}$
9.	Improve($\mathcal{S}_i \cup \mathcal{S}_j, \mathcal{R}_i \cup \mathcal{R}_j$)
10.	$\text{csum}=0, \text{isum}=0$
11.	For all $s_i \in \mathcal{S}_i \cup \mathcal{S}_j$,
12.	$\text{csum} += C_{\text{NC-CC}}^{w_i}(s_i, \mathcal{S}_i \cup \mathcal{S}_j, \mathcal{R}_i \cup \mathcal{R}_j, d_i)$
13.	$\text{isum} += C_{\text{NC-CC}}^{w_i}(s_i, \mathcal{S}_i, \mathcal{R}_i, d_i) + C_{\text{NC-CC}}^{w_i}(s_i, \mathcal{S}_j, \mathcal{R}_j, d_i)$
14.	If $\text{csum} > \text{isum}$
15.	Add $(\mathcal{S}_i \cup \mathcal{S}_j, \mathcal{R}_i \cup \mathcal{R}_j)$ to $\mathcal{L}_{\text{temp}}$
16.	$\text{merge_count}++$
17.	else
18.	If $(\mathcal{S}_i, \mathcal{R}_i) \notin \mathcal{L}_{\text{temp}}$, add $(\mathcal{S}_i, \mathcal{R}_i)$ to $\mathcal{L}_{\text{temp}}$
19.	If $(\mathcal{S}_j, \mathcal{R}_j) \notin \mathcal{L}_{\text{temp}}$, add $(\mathcal{S}_j, \mathcal{R}_j)$ to $\mathcal{L}_{\text{temp}}$
20.	If($\text{merge_count} > 0$)
21.	Set $\mathcal{L}_{k+1} = \emptyset$
22.	For each $(\mathcal{S}_i, \mathcal{R}_i)$ pair in $\mathcal{L}_{\text{temp}}$,
23.	If \mathcal{S}_i was constructed by combining two session groups in \mathcal{L}_{k-1} , then
24.	Set $\text{gain}_i = \sum_{s_j \in \mathcal{S}_i} C_{\text{NC-CC}}^{w_i}(s_j, \mathcal{S}_i, \mathcal{R}_i, d_j) - C_{\text{D}}^{w_i}(s_j, d_j)$
25.	Else, set $\text{gain}_i = 0$
26.	Set $\mathcal{L}_{\text{singles}} = \emptyset$
27.	Consider all $(\mathcal{S}_i, \mathcal{R}_i)$ pairs in $\mathcal{L}_{\text{temp}}$ in the decreasing order of gain_i values. For each $(\mathcal{S}_i, \mathcal{R}_i)$
28.	If $(\mathcal{S}_i, \mathcal{R}_i)$ was constructed by combining two session groups from \mathcal{L}_{k-1}
29.	Denote the two session and relay groups by $(\mathcal{S}_i^1, \mathcal{R}_i^1)$ and $(\mathcal{S}_i^2, \mathcal{R}_i^2)$
30.	If neither \mathcal{S}_i^1 nor \mathcal{S}_i^2 is part of some $(\mathcal{S}_j, \mathcal{R}_j)$ that is already in \mathcal{L}_{k+1}
31.	Then, add $(\mathcal{S}_i, \mathcal{R}_i)$ to \mathcal{L}_{k+1}
32.	Else, among $(\mathcal{S}_i^1, \mathcal{R}_i^1)$ and $(\mathcal{S}_i^2, \mathcal{R}_i^2)$, add the pair, whose session group is not already in \mathcal{L}_{k+1} , to $\mathcal{L}_{\text{singles}}$
33.	Else, if \mathcal{S}_i is not part of some $(\mathcal{S}_j, \mathcal{R}_j)$ already in \mathcal{L}_{k+1}
34.	Then, add $(\mathcal{S}_i, \mathcal{R}_i)$ to \mathcal{L}_{k+1}
35.	If $\mathcal{L}_{\text{singles}} \neq \emptyset$
36.	Then, for all $(\mathcal{S}_i, \mathcal{R}_i)$ in $\mathcal{L}_{\text{singles}}$
37.	If \mathcal{S}_i is not part of some $(\mathcal{S}_j, \mathcal{R}_j)$ already in \mathcal{L}_{k+1} ,
38.	Add $(\mathcal{S}_i, \mathcal{R}_i)$ to \mathcal{L}_{k+1}
39.	$k++$
40.	Goto Step 7
41.	Else, if ($\text{merge_count} == 0$)
42.	Then none of the matching pairs in \mathcal{L}_k were combined, so we will stop
43.	List \mathcal{L}_k contains the grouping and matching solution
Improve($\mathcal{S}_i \cup \mathcal{S}_j, \mathcal{R}_i \cup \mathcal{R}_j$)	
1.	For all $r_k \in (\mathcal{R}_i \cup \mathcal{R}_j) - (\mathcal{R}_i \cap \mathcal{R}_j)$
2.	If $\sum_{s_m \in \mathcal{S}_i \cup \mathcal{S}_j} C_{\text{NC-CC}}^{w_i}(s_j, \mathcal{S}_i \cup \mathcal{S}_j, \mathcal{R}_i \cup \mathcal{R}_j - r_k, d_j) > \sum_{s_m \in \mathcal{S}_i \cup \mathcal{S}_j} C_{\text{NC-CC}}^{w_i}(s_m, \mathcal{S}_i \cup \mathcal{S}_j, \mathcal{R}_i \cup \mathcal{R}_j, d_m)$
3.	If $r_k \in \mathcal{R}_i$, then $\mathcal{R}_i \leftarrow \mathcal{R}_i - r_k$
4.	If $r_k \in \mathcal{R}_j$, then $\mathcal{R}_j \leftarrow \mathcal{R}_j - r_k$

Fig. 10. Pseudocode of the G²M main algorithm.

- For Entry 6 in Table 5, the proposed merger is not desirable. Further, match $(\mathcal{S}_3, \mathcal{R}_3)$ prior to this proposed merger was already added to $\mathcal{L}_{\text{temp}}$ earlier under Entry 3. So, $(\mathcal{S}_3, \mathcal{R}_3)$ is dropped in Entry 6 of $\mathcal{L}_{\text{temp}}$. For the same reason, for Entry 7 in Table 5, the match $(\mathcal{S}_1, \mathcal{R}_1)$ was already included and will not be added again to $\mathcal{L}_{\text{temp}}$ in Table 5.

Table 5

List $\mathcal{L}_{\text{temp}}$.

	Merged pair	Temporary gain
1	$\{(s_0, d_0), (s_1, d_1)\}, \{r_0, r_1, r_2\}$	15
2	$\{(s_0, d_0), (s_2, d_2)\}, \{r_1, r_3\}$	10
3*	$\{(s_0, d_0)\}, \{r_0, r_1\}$	0
	$\{(s_3, d_3)\}, \{\emptyset\}$	0
4	$\{(s_0, d_0), (s_4, d_4)\}, \{r_0, r_1, r_4\}$	4
5	$\{(s_1, d_1), (s_2, d_2)\}, \{r_1, r_3\}$	7
6*	$\{(s_1, d_1)\}, \{r_1, r_2, r_3\}$	0
7*	$\{(s_4, d_4)\}, \{r_1, r_4\}$	0
8	$\{(s_2, d_2), (s_3, d_3)\}, \{r_3\}$	1
9	$\{(s_2, d_2), (s_4, d_4)\}, \{r_1\}$	6
10	$\{(s_3, d_3), (s_4, d_4)\}, \{r_4\}$	3

Table 6

List \mathcal{L}_2 .

	Merged sessions	Merged relays	Increase in sum of weighted rates
	$\{(s_0, d_0), (s_1, d_1)\}$	$\{r_0, r_1, r_2\}$	15
	$\{(s_2, d_2), (s_4, d_4)\}$	$\{r_1\}$	6
	$\{(s_3, d_3)\}$	$\{\emptyset\}$	0

Table 5 shows the final $\mathcal{L}_{\text{temp}}$. To create \mathcal{L}_2 , we consider entries in $\mathcal{L}_{\text{temp}}$ in Table 5 in the order of decreasing value of temporary gains.

- We first consider the entry with the largest gain 15, i.e., Entry 1, and add it to \mathcal{L}_2 , which is empty initially.
- We now consider Entry 2, which offers the second largest temporary gain of 10. This proposed match includes \mathcal{S}_0 and \mathcal{S}_2 . Since we have already added \mathcal{S}_0 to \mathcal{L}_2 , we cannot do it again. But $\mathcal{S}_2 = \{(s_2, d_2)\}$ is still not in \mathcal{L}_2 . So, we will replace entry 2 in $\mathcal{L}_{\text{temp}}$ with $(\mathcal{S}_2, \mathcal{R}_2)$ having a zero temporary gain. This will effectively make this entry to be considered last.
- The next entry to be considered is Entry 5 with a temporary gain of 7. The session groups in Entry 5 include \mathcal{S}_1 and \mathcal{S}_2 . Again, since \mathcal{S}_1 is already in \mathcal{L}_2 , we cannot add it again. For \mathcal{S}_2 , it has already been considered earlier for Entry 2. So, Entry 5 can be deleted as all of its sessions have already been considered in earlier steps.
- The next entry is Entry 9 with a temporary gain of 6. As none of these sessions are in \mathcal{L}_2 , it will be added to \mathcal{L}_2 .
- This process continues until all sessions are included in \mathcal{L}_2 , as shown in Table 6.

Future iterations (to obtain $\mathcal{L}_3, \mathcal{L}_4, \dots$) follow the same token. This procedure will terminate when we can no longer make improvement on the current matching list.

5.4. Runtime complexity

Initialization of \mathcal{L}_1 . To create the initial list of matchings for session s_i , we need to sort relays for s_i in increasing order of $\text{SNR}_{s_i r_j}$ values. The complexity of sorting relays is $O(|\mathcal{R}| \log_2 |\mathcal{R}|)$ per session. Thus, the complexity of sorting for all the sessions is $O(|\mathcal{S}| \cdot |\mathcal{R}| \cdot \log_2 |\mathcal{R}|)$.

Afterwards, for each session, there are at most $|\mathcal{R}|$ iterations to get the initial matching, with each iteration having a worst case complexity of $O(|\mathcal{R}|)$. Thus, the complexity for all sessions is $O(|\mathcal{S}| \cdot |\mathcal{R}|^2)$.

Thus, the worst case complexity to create initial matchings for all $|\mathcal{S}|$ sessions is $O(|\mathcal{S}| \cdot |\mathcal{R}| \cdot \log_2 |\mathcal{R}|) + O(|\mathcal{S}| \cdot |\mathcal{R}|^2) = O(|\mathcal{S}| \cdot |\mathcal{R}|^2)$.

Creation of $\mathcal{L}_{\text{temp}}$. Next, in the main program of the algorithm, every iteration has to go through $O(|\mathcal{S}|^2)$ matching pairs to check

for possible merger, and create the list $\mathcal{L}_{\text{temp}}$. For every matching pair in the iteration, a check for possible merger involves:

- Merging the corresponding relay groups. The complexity of this step is $O(|\mathcal{R}|)$.
- If the merger improves the objective value, then the merged pair is added to $\mathcal{L}_{\text{temp}}$ and requires $O(1)$ complexity.
- In case the pair of matchings cannot be merged, then we may have to add individual matchings to the list $\mathcal{L}_{\text{temp}}$. This will require $O(|\mathcal{S}|^2)$ checks within $\mathcal{L}_{\text{temp}}$ to see whether or not the individual sessions already exist in $\mathcal{L}_{\text{temp}}$.

So, the worst case complexity for creating list $\mathcal{L}_{\text{temp}}$ is $O(|\mathcal{S}|^2 \cdot (|\mathcal{R}| + |\mathcal{S}|^2)) = O(|\mathcal{S}|^2 \cdot |\mathcal{R}| + |\mathcal{S}|^4)$.

Sorting of $\mathcal{L}_{\text{temp}}$. Next, we need to sort the list $\mathcal{L}_{\text{temp}}$. This will have a worst case complexity of $O(|\mathcal{S}|^2 \log_2 |\mathcal{S}|^2) = O(|\mathcal{S}|^2 \log_2 |\mathcal{S}|)$.

Creation of list for next iteration. To create the next list of matching from $\mathcal{L}_{\text{temp}}$, we need to go through all $O(|\mathcal{S}|^2)$ session groups in the list $\mathcal{L}_{\text{temp}}$ in increasing order of the temporary gain values. Before moving a matching from $\mathcal{L}_{\text{temp}}$ to the new list for next iteration, we need to check that we are not adding a session group multiple times in the new list. Checking this will incur an overhead of $O(|\mathcal{S}|)$ per matching. Thus, the total worst case complexity to create the list for next iteration is $O(|\mathcal{S}|^2 \cdot |\mathcal{S}|) = O(|\mathcal{S}|^3)$.

Finally, as we start from a list of $|\mathcal{S}|$ matchings, and every improving iteration merge at least two session groups, there are no more than $O(|\mathcal{S}|)$ iterations in the worst case. Thus, the total complexity of G²M is $O(|\mathcal{S}| \cdot [|\mathcal{S}| \cdot |\mathcal{R}|^2 + |\mathcal{S}|^2 |\mathcal{R}| + |\mathcal{S}|^4 + |\mathcal{S}|^2 \log |\mathcal{S}| + |\mathcal{S}|^3]) = O(|\mathcal{S}|^2 |\mathcal{R}|^2 + |\mathcal{S}|^3 |\mathcal{R}| + |\mathcal{S}|^5)$.

5.5. Further improvements to G²M

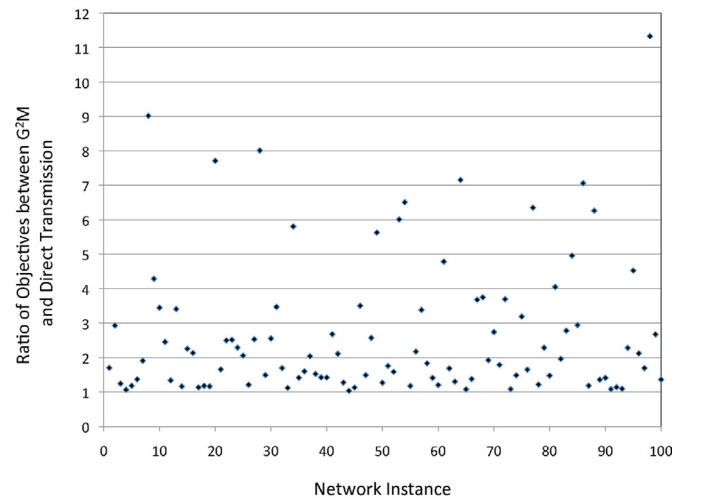
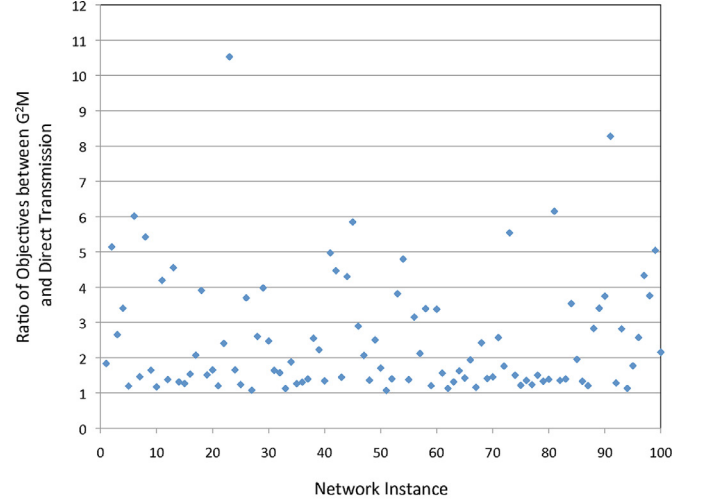
Many improvements to the baseline G²M algorithm are possible. In this section, we discuss two possible improvements.

Recall that in the main program of G²M, we only consider possible merger of two groups of sessions at a time. It may also be beneficial to consider three or more groups of sessions at the same time. This is because that although proposed merger of two groups of sessions may not improve the objective value, merging of three or more groups of sessions may lead to an improvement. As an example, suppose that the proposed merger of matchings $(\mathcal{S}_0, \mathcal{R}_0)$ and $(\mathcal{S}_1, \mathcal{R}_1)$ will reduce the objective value, and the same situation will occur if we merge $(\mathcal{S}_1, \mathcal{R}_1)$ and $(\mathcal{S}_2, \mathcal{R}_2)$, or $(\mathcal{S}_0, \mathcal{R}_0)$ and $(\mathcal{S}_2, \mathcal{R}_2)$. But if we merge $(\mathcal{S}_0, \mathcal{R}_0)$, $(\mathcal{S}_1, \mathcal{R}_1)$, and $(\mathcal{S}_2, \mathcal{R}_2)$ all together, it may improve the objective value. This feature (i.e., considering merger of three or more sessions groups) can be easily incorporated into the main program of G²M.

Another improvement that we can make to G²M is to add more dynamics in merging during each iteration. In the baseline algorithm, once we merge two session groups, they will never be separated again, which may limit other opportunities. That is, a “match-and-never-separate” approach is somewhat rigid and may miss an optimal solution. As an improvement, we may want to check whether the objective can be improved by removing some sessions from the merged group. This can be done by incorporating a second phase in the main program.

6. Numerical results

In this section, we present numerical results to demonstrate the performance and efficiency of our G²M algorithm. Our goals are threefold: (i) to show G²M algorithm offers better results than direct transmission, (ii) to demonstrate that the solutions constructed by G²M are close to the optimal solutions obtained by CPLEX solver [7], and (iii) to show that the running time of G²M is orders of magnitude lower than CPLEX solver.



(a) $w_i = 1$ for $i = 0, \dots, |\mathcal{S}| - 1$.

(b) Each w_i is randomly generated between $[0, 1]$; $i = 0, \dots, |\mathcal{S}| - 1$.

Fig. 11. Ratios between the objective values under G²M and direct transmission.

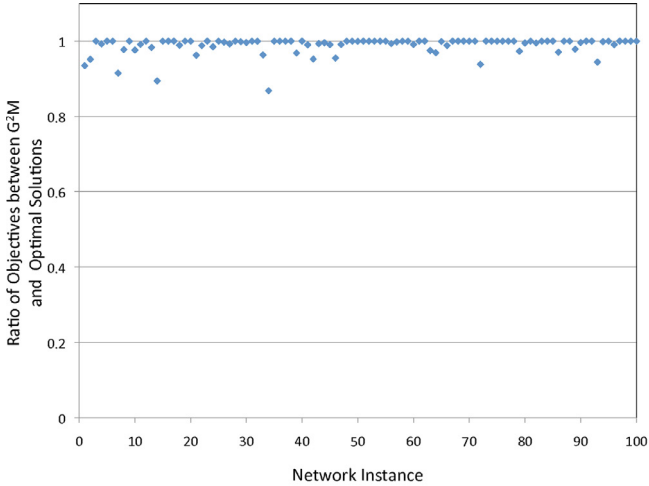
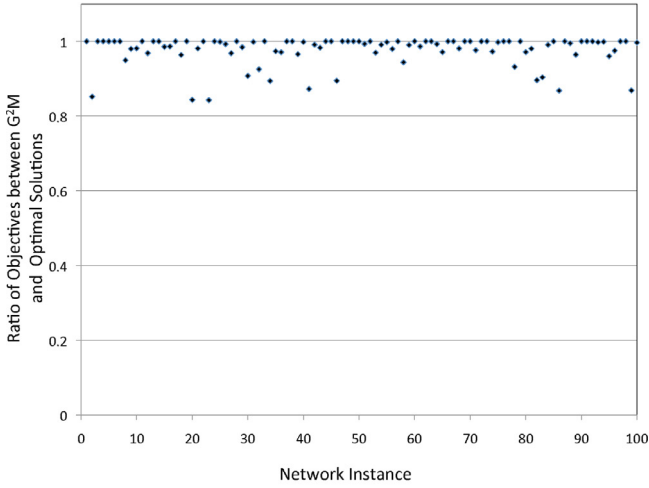
6.1. Parameter settings

For all network instances used in this simulation study, we assume the transmission power at each node to be 1 W. The available transmission bandwidth at every node is 20 MHz, and the variance of white Gaussian background noise at all nodes is 10^{-10} W. The channel gain between two nodes s and d is modeled as $|h_{sd}|^2 = ||s - d||^{-4}$, where $||s - d||$ is the distance between s and d (in meters).

6.2. Results

6.2.1. G²M vs. direct transmission

We consider 100 different randomly generated network instances, each with 30 nodes (7 source-destination pairs and 16 relay nodes). For each instance, the nodes are randomly deployed in an area of size 1200 m \times 1200 m square. We calculate the objective value for each network instance under both G²M and direct transmission. Fig. 11a plots the ratio of the objective values obtained under G²M and those under direct transmission when all the weights in the network are set to 1. Similarly, Fig. 11b plots the ratio when each session's weight in the network is randomly chosen between 0 and 1. In Fig. 11a, the average ratio is 2.53 (with a variance of 2.83); in Fig. 11b, the average ratio is 2.67 (with a vari-

(a) $w_i = 1$ for $i = 0, \dots, |\mathcal{S}| - 1$.(b) Each w_i is randomly generated between $[0, 1]$; $i = 0, \dots, |\mathcal{S}| - 1$.**Fig. 12.** Ratio between the objective values under G²M and CPLEX.

ance of 3.98). Note that under any network instance in each figure, the ratio between the two is no less than 1, due to our design of the G²M algorithm.

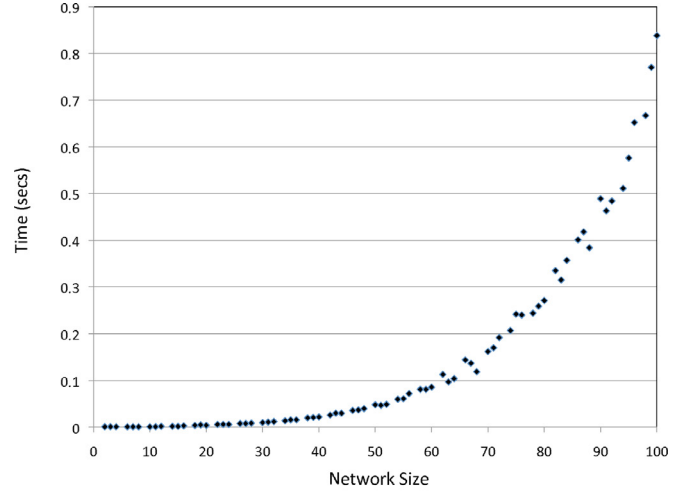
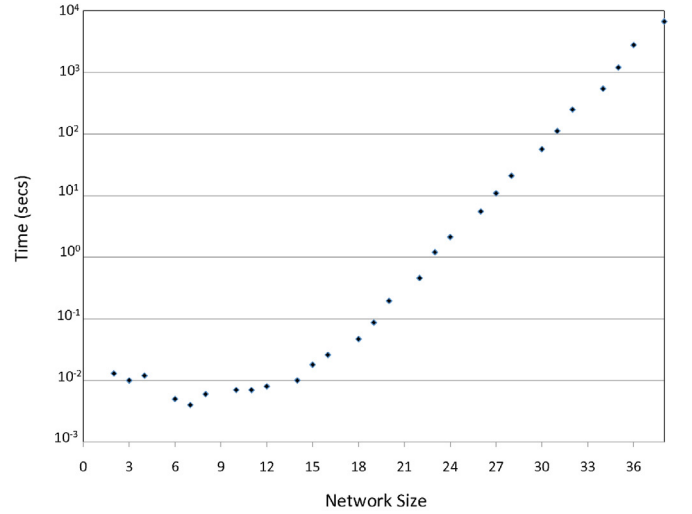
6.2.2. Near-Optimality of G²M

To validate the performance of G²M, we compare the results by G²M to the optimal solutions obtained by solving a mathematical formulation of our session/relay grouping and matching problem. A mathematical formulation of this problem is given in the Appendix, which is in the form of 0–1 integer linear program (ILP).

Fig. 12a shows the ratio between the objective values obtained by G²M over those from CPLEX when the weight of each session is set to 1. Similarly, Fig. 12b shows the ratio between the two when the weight of each session is randomly set between $[0, 1]$. As we can see, the performance of G²M is highly competitive in both cases. It is 98.8% of optimal on average (with a variance of 0.05) for fixed weights, and 97.7% optimal on average (with a variance of 0.15) for random weights.

6.2.3. Time complexity

We now compare the time complexity between the G²M algorithm and the CPLEX solver. Due to its polynomial complexity, we find that the running time of G²M is orders of magnitude faster than CPLEX, which has an exponential time complexity to solve an integer linear program. As an illustration, we start with a network

(a) Running time of G²M for network size between 2 to 100.

(b) Running time of CPLEX solver for network size between 2 to 38.

Fig. 13. Comparison of running time between G²M and CPLEX. (Note the difference in scales on both x- and y-axis in the two figures.)

having only one session, and add more relays and sessions to grow the network size. For each network instance, we compare the time required to find the solution by G²M and that by the CPLEX solver. Fig. 13a shows that the time required to get the solution under G²M when the network size grows from 2 to 100. On the other hand, Fig. 13b shows the time required to obtain optimal solutions from CPLEX when the network size grows from 2 to 38. Note that the y-axis in Fig. 13b is in log scale rather than the linear scale in Fig. 13a.

7. Related work

Although CC has been an active research area for many years (see, e.g. [1,5,8,9,12–15,17,28,31,37]), recent advances in employing NC in CC (so-called NC-CC) has created a new research paradigm for the communication network community. To date, research on NC-CC is still in its early stage and results remain very limited [4,19,25,33,36,38]. In [4], Bao and Li were the first to employ NC-CC in a multi-source single-destination network. Their focus was on developing coding mechanisms that could be used by the source nodes to cooperate with each other. In [19], Peng et al. considered a network with a single relay node and multiple source-destination pairs, and studied the outage probability of the entire

network when NC-CC is employed. Sharma et al. [25] also considered a network with a single relay node and multiple source-destination pairs, and derived the data rate equations for individual sessions under NC-CC. Xiao et al. [33] considered a two-source single-destination network and showed that NC can help CC reduce packet error rates. In [36] and [38], the NC-CC framework was limited in exploiting NC only in case of *bi-directional* traffic and by using a *single* relay node. We have shown in this paper that NC-CC is beneficial in unidirectional traffic as well, and multiple relay NC-CC can be significantly better than single relay NC-CC. As a result, limiting the work to bi-directional traffic only and the use of a single relay node limits the potential gains of their approach in an ad-hoc network. In fact, a common limitation of all these prior efforts is the use of only a single relay node. As a result, they could not benefit from any performance gains that can be offered by multiple relay nodes. NC-CC with multiple relay nodes was first explored by Sharma et al. very recently [27], where they showed that a proper choice of a group of relay nodes could have a significant impact on NC-CC's performance. However, the problem on how to group sessions, relay nodes, and match them together remains open. This paper is the first attempt to address this important problem.

8. Conclusion

In this paper, we studied how NC-CC behaves when multiple relay nodes are employed. We derived closed form equations for mutual information and achievable data rate for each session under such setting. Our derivation was based on generalizing the three-node relay channel model with multiple relays and then taking consideration of multiple sessions. Our results in this paper offer an important building block on the theory of NC-CC. To demonstrate the application of our theoretical result, we solved a joint optimization problem in multiple dimensions, such as session grouping, relay node grouping, and matching of session/relay groups. After showing that this problem was NP-hard, we presented a polynomial time heuristic algorithm to this problem. Using simulation results, we showed that the proposed algorithm is highly competitive and efficient: it is able to offer a performance close to a centralized solver while being orders of magnitude faster.

Acknowledgments

This work was supported in part by NSF Grant 1064953 and ONR Grant N00014-15-1-2926. The work of S. Kompella was supported in part by the ONR. The authors thank Virginia Tech's Advanced Research Computing for giving them access to the BlueRidge computer cluster.

Appendix

A mathematical formulation of session/relay grouping and matching problem. A *feasible* session group (with respect to a relay group) is defined as a group of sessions sharing the relay group under NC-CC where the weighted sum of session rates in the group is not less than the weighted sum of session rates under direct transmission. So, we denote $\mathcal{I}_{\mathcal{R}_j}$ as the set of all feasible session groups for the relay group $\mathcal{R}_j \in \mathcal{R}$. The size of the set $\mathcal{I}_{\mathcal{R}_j}$ can be exponential in the worst case. We further denote $\mathcal{J}_{\mathcal{R}_j}^{s_i} \subset \mathcal{I}_{\mathcal{R}_j}$ as the set of feasible session groups that contain the source node s_i . Thus, we have $\bigcup_{s_i \in \mathcal{S}} \mathcal{J}_{\mathcal{R}_j}^{s_i} = \mathcal{I}_{\mathcal{R}_j}$.

We use the notation $\mathcal{G} \in \mathcal{S}$ to denote some group of sessions. We define a binary variable $X_{\mathcal{R}_j}^{\mathcal{G}}$ as follows:

$$X_{\mathcal{R}_j}^{\mathcal{G}} = \begin{cases} 1 & \text{if session group } \mathcal{G} \text{ uses relay group } \mathcal{R}_j \text{ for NC-CC, } \mathcal{G} \in \mathcal{I}_{\mathcal{R}_j}, \\ 0 & \text{otherwise.} \end{cases}$$

Since a source node s_j can be in at most one session group, we have:

$$\sum_{\mathcal{R}_j \in \mathcal{N}_r} \sum_{\mathcal{G} \in \mathcal{J}_{\mathcal{R}_j}^{s_i}} X_{\mathcal{R}_j}^{\mathcal{G}} \leq 1 \quad (s_i \in \mathcal{S}), \quad (22)$$

where \mathcal{N}_r is the set of relay groups (among all possible relay groups) that are feasible for at least one session group.

Note that by (22), a session can belong to only one session group, and there exists $|\mathcal{S}|$ source nodes. Thus, the number of groups that may be ultimately formed associated with any of the relay groups is at most $|\mathcal{S}|$. Further, note that (22) is not prohibiting relay nodes from participating in multiple relay groups, neither is it prohibiting multiple session groups to use same relay groups.

For the objective of maximizing the sum of weighted rates among all source nodes in the network, we can formulate the session/relay grouping and matching problem as follows:

$$\begin{aligned} \text{Maximize } & \sum_{s_i \in \mathcal{S}} \left[\left(\sum_{\mathcal{R}_j \in \mathcal{N}_r} \sum_{\mathcal{G} \in \mathcal{J}_{\mathcal{R}_j}^{s_i}} X_{\mathcal{R}_j}^{\mathcal{G}} \cdot w_i \cdot C_{\text{NC-CC}}(s_i, \mathcal{G}, \mathcal{R}_j, d_i) \right) \right. \\ & \left. + \left(1 - \sum_{\mathcal{R}_j \in \mathcal{N}_r} \sum_{\mathcal{G} \in \mathcal{J}_{\mathcal{R}_j}^{s_i}} X_{\mathcal{R}_j}^{\mathcal{G}} \right) \cdot w_i \cdot C_D(s_i, d_i) \right] \\ \text{subject to } & \sum_{\mathcal{R}_j \in \mathcal{N}_r} \sum_{\mathcal{G} \in \mathcal{J}_{\mathcal{R}_j}^{s_i}} X_{\mathcal{R}_j}^{\mathcal{G}} \leq 1 \quad (s_i \in \mathcal{S}) \\ & X_{\mathcal{R}_i}^{\mathcal{G}} \in \{0, 1\}, \quad (\mathcal{R}_i \in \mathcal{R}, \mathcal{G} \in \mathcal{I}_{\mathcal{R}_i}). \end{aligned}$$

Note that the objective function contains the sum of two different terms for every source node s_i in the network. The first term is the achievable rate under NC-CC while the second term contains the achievable rate under direct transmission; only one of these two terms will be non-zero and taken into the summation over $s_i \in \mathcal{S}$.

The above optimization problem is a 0–1 integer linear programming (ILP) problem, with an exponential number of variables ($X_{\mathcal{R}_j}^{\mathcal{G}}$) in the worst case. Due to the combinatorial nature of the problem, any alternative formulation will also involve integer (or binary) variables in it.

References

- [1] B. Aazhang, R. S. Blum, J. N. Laneman, K. J. R. Liu, W. Su, A. Wittneben (Eds.), Special issue on cooperative communications and networking, IEEE J. Sel. Areas Commun. vol. 25, 2007.
- [2] A. Antonopoulos, C. Verikoukis, C. Skianis, O.B. Akan, Energy efficient network coding-based MAC for cooperative ARQ wireless networks, Elsevier Ad Hoc Netw. 11 (1) (2013) 190–200.
- [3] A. Antonopoulos, A.S. Lalos, M.D. Renzo, C. Verikoukis, Cross-layer theoretical analysis of NC-aided cooperative ARQ protocols in correlated shadowed environments, IEEE Trans. Veh. Technol. 64 (9) (2015) 4074–4087.
- [4] X. Bao, J. Li, Adaptive network coded cooperation (ANCC) for wireless relay networks: matching code-on-graph with network-on-graph, IEEE Trans. Wireless Commun. 7 (2) (2008) 574–583.
- [5] T.M. Cover, A.E. Gamal, Capacity theorems for the relay channel, IEEE Trans. Inf. Theory 25 (5) (1979) 572–584.
- [6] F.H.P. Fitzek, M. Dohler, I. Opperman, A. Molinaro, M. Katz, C. Bettstetter (Eds.), Special issue on cooperation in wireless networks, J. Wirel. Pers. Commun. vol. 43, 2007.
- [7] IBM ILOG CPLEX. <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html> (accessed 14.09.16).
- [8] G. Jakllari, S.V. Krishnamurthy, M. Faloutsos, P.V. Krishnamurthy, O. Ercetin, A cross-layer framework for exploiting virtual MISO links in mobile ad hoc networks, IEEE Trans. Mobile Comput. 6 (5) (2007) 579–594.

- [9] G. Jakllari, S.V. Krishnamurthy, M. Faloutsos, P.V. Krishnamurthy, On broadcasting with cooperative diversity in multi-hop wireless networks, *IEEE J. Sel. Areas Commun.* 25 (2) (2007) 484–496.
- [10] S. Katti, S. Gollakotta, D. Katabi, Embracing wireless interference: Analog network coding, in: *Proceedings of the ACM SIGCOMM*, August 27–31, Kyoto, Japan, 2007, pp. 397–408.
- [11] G. Kramer, I. Maric, R.D. Yates, “Cooperative communications,” *Foundations and Trends in Networking*, 1, Now Publishers, 2007.
- [12] G. Kramer, I. Maric, R.D. Yates, Cooperative communications, *Foundations and Trends in Networking*, 1, Now Publishers, 2006.
- [13] G. Kramer, R. Berry, A.E. Gamal, H.E. Gamal, M. Franceschetti, M. Gastpar, J.N. Laneman (Eds.), Special issue on models, theory, and codes for relaying and cooperation in communication networks, *IEEE Trans. Inf. Theory*, vol. 53, 2007.
- [14] S. Lakshmanan, R. Sivakumar, Diversity routing for multi-hop wireless networks with cooperative transmissions, in: *Proceedings of IEEE SECON*, Rome, Italy, June 22–26, 2009.
- [15] J.N. Laneman, D.N.C. Tse, G.W. Wornell, Cooperative diversity in wireless networks: efficient protocols and outage behavior, *IEEE Trans. Inf. Theory* 50 (12) (2004) 3062–3080.
- [16] L.E. Li, R. Alimi, D. Shen, H. Vishwanathan, Y.R. Yang, A general algorithm for interference alignment and cancellation in wireless networks, in: *Proceedings of IEEE INFOCOM*, San Diego, CA March 15–19, 2010.
- [17] K.J.R. Liu, A.K. Sadek, W. Su, A. Kwasinski, Cooperative Communications and Networking, Cambridge University Press, 2009.
- [18] K. Ntontin, M.D. Renzo, A.I. Perez-Neira, C. Verikoukis, Analog network coding in the multiple access relay channel: Error rate analysis and optimal power allocation, *IEEE Trans. Wirel. Commun.* 14 (6) (2015) 3015–3032.
- [19] C. Peng, Q. Zhang, M. Zhao, Y. Yao, On the performance analysis of network-coded cooperation in wireless networks, in: *Proceedings of IEEE INFOCOM*, May 6–12, Anchorage, Alaska, 2007, pp. 1460–1468.
- [20] J.L. Rebelatto, B.F. Uchoa-Filho, Y. Li, B. Vucetic, Multi-user cooperative diversity through network coding based on classical coding theory, *IEEE Trans. Signal Process.* 60 (2) (2011) 916–926.
- [21] M.D. Renzo, M. Iezzi, F. Graziosi, On diversity order and coding gain of multi-source multirelay cooperative wireless networks with binary network coding, *IEEE Trans. Veh. Technol.* 63 (3) (2013) 1138–1157.
- [22] M.D. Renzo, M. Iezzi, F. Graziosi, Error performance and diversity analysis of multi-source multi-relay wireless networks with binary network coding and cooperative MRC, *IEEE Trans. Wirel. Commun.* 12 (6) (2013) 2883–2903.
- [23] M.D. Renzo, On the achievable diversity of repetition-based and relay selection network-coded cooperation, *IEEE Trans. Commun.* 62 (7) (2014) 2296–2313.
- [24] I. Safak, E. Aktas, A.O. Yilmaz, Error rate analysis of GF(q) network coded detect-and-forward wireless relay networks using equivalent relay channel models, *IEEE Trans. Wirel. Commun.* 12 (8) (2013) 3908–3919.
- [25] S. Sharma, Y. Shi, J. Liu, Y.T. Hou, Is network coding always good for cooperative communications, in: *Proceedings of IEEE INFOCOM*, San Diego, CA, March 15–19, 2010.
- [26] S. Sharma, Y. Shi, Y.T. Hou, H.D. Sherali, S. Kompella, Optimizing network-coded cooperative communications via joint session grouping and relay node selection, in: *Proceedings of IEEE INFOCOM*, Shanghai, China, April 10–15, 2011.
- [27] S. Sharma, Y. Shi, Y.T. Hou, S. Kompella, Achievable rate analysis in network-coded cooperative communications with multiple relay nodes, *Proceeding of IEEE ICC – Wireless Networking Symposium*, Kyoto, Japan, June 5–9, 2011.
- [28] S. Sharma, Y. Shi, Y.T. Hou, S. Kompella, An optimal algorithm for relay node assignment in cooperative ad hoc networks, *IEEE/ACM Trans. Netw.* 19 (1) (2011) 879–892.
- [29] H. Topakkaya, Z. Wang, Wireless network code design and performance analysis using diversity-multiplexing tradeoff, *IEEE Trans. Commun.* 59 (2) (2011) 488–496.
- [30] Y. Tu, G.J. Pottie, Coherent cooperative transmission from multiple adjacent antennas to a distant stationary antenna through AWGN channels, in: *Proceedings of IEEE Vehicular Technology Conference*, May 6–9, Birmingham, AL, 2002, pp. 130–134.
- [31] E.C. van der Meulen, Three terminal communication channels, *Adv. Appl. Probab.* 3 (1971) 120–154.
- [32] X. Wang, J. Li, Network coding aware cooperative MAC protocol for wireless ad hoc networks, *IEEE Trans. Parallel Distrib. Syst.* 25 (1) (2014) 167–179.
- [33] L. Xiao, T.E. Fuja, J. Kliewer, D.J. Costello, A network coding approach to cooperative diversity, *IEEE Trans. Inf. Theory* 53 (10) (2007) 3714–3722.
- [34] M. Xiao, M. Skoglund, Multiple-user cooperative communications based on linear network coding, *IEEE Trans. Commun.* 58 (12) (2010) 3345–3351.
- [35] M. Xiao, J. Kliewer, M. Skoglund, Design of network codes for multiple-user multiple relay wireless networks, *IEEE Trans. Commun.* 60 (12) (2012) 3755–3766.
- [36] H. Xu, B. Li, XOR-assisted cooperative diversity in OFDMA wireless networks: Optimization framework and approximation algorithms, in: *Proceedings of IEEE INFOCOM*, April 19–25, Rio de Janeiro, Brazil, 2009, pp. 2141–2149.
- [37] E.M. Yeh, R.A. Berry, Throughput optimal control of cooperative relay networks, *IEEE Trans. Inf. Theory* 53 (10) (2007) 3827–3833.
- [38] J. Zhang, Q. Zhang, Cooperative network coding-aware routing for multi-rate wireless networks, in: *Proceedings of IEEE INFOCOM*, April 19–25, Rio de Janeiro, Brazil, 2009, pp. 181–189.



Sushant Sharma received his Ph.D. degree in computer science from Virginia Tech, Blacksburg, VA, in 2010. He is currently working as a Software Engineer in Microsoft Azure, Redmond, WA. His research interests include developing novel algorithms to solve network optimization problems, and building distributed networked systems.



Yi Shi received his Ph.D. degree in Computer Engineering from Virginia Tech, Blacksburg, VA in 2007. He is currently a Senior Research Scientist at Intelligent Automation Inc., Rockville, MD, and an Adjunct Assistant Professor at Virginia Tech. His research focuses on optimization and algorithm design for wireless networks. He was a recipient of IEEE INFOCOM 2008 Best Paper Award and the only Best Paper Award Runner-Up of IEEE INFOCOM 2011.



Y. Thomas Hou is the Bradley Distinguished Professor of Electrical and Computer Engineering at Virginia Tech, Blacksburg, VA. He received his Ph.D. degree from NYU Polytechnic School of Engineering (formerly Polytechnic University). His current research focuses on developing innovative solutions to complex cross-layer optimization problems in wireless and mobile networks. He has published two graduate textbooks: *Applied Optimization Methods for Wireless Networks* (Cambridge University Press, 2014) and *Cognitive Radio Communications and Networks: Principles and Practices* (Academic Press/Elsevier, 2009). He is the Steering Committee Chair of IEEE INFOCOM conference and a member of the IEEE Communications Society Board of Governors.



Sastry Kompella received his Ph.D. degree in computer engineering from Virginia Tech, Blacksburg, Virginia, in 2006. Currently, he is the head of Wireless Network Theory section, Information Technology Division at the U.S. Naval Research Laboratory (NRL), Washington, DC. His research focuses on complex problems in cross-layer optimization and scheduling in wireless and cognitive radio networks.



Scott F. Midkiff is Professor & Vice President for Information Technology and Chief Information Officer at Virginia Tech, Blacksburg, VA. From 2009 to 2012, Prof. Midkiff was the Department Head of the Bradley Department of Electrical and Computer Engineering at Virginia Tech. From 2006 to 2009, he served as a program director at the National Science Foundation. Prof. Midkiff's research interests include wireless and ad hoc networks, network services for pervasive computing, and cyber-physical systems.